

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 June 2001 (07.06.2001)

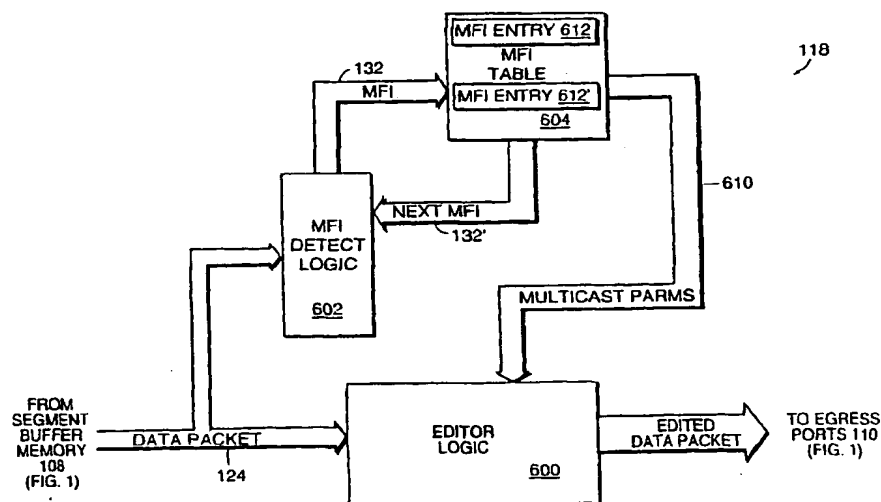
PCT

(10) International Publication Number  
**WO 01/41364 A2**

- (51) International Patent Classification<sup>7</sup>: **H04L 12/18**, 29/06, 12/46
- (21) International Application Number: **PCT/CA00/01420**
- (22) International Filing Date:  
30 November 2000 (30.11.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
09/453,344 1 December 1999 (01.12.1999) US
- (71) Applicant (for all designated States except US): **MOSAID TECHNOLOGIES INCORPORATED** [CA/CA]; 11 Hines Road, Kanata, Ontario K2K 2X1 (CA).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **BROWN, David, A.** [CA/CA]; 110 Allissia Crescent, Carp, Ontario K0A 1L0 (CA).
- (74) Agent: **PILLAY, Kevin**; Fasken Martineau DuMoulin LLP, Suite 4200, Box 20, Toronto-Dominion Centre, Toronto Dominion Bank Tower, Toronto, Ontario M5K 1N6 (CA).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— Without international search report and to be republished upon receipt of that report.

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR WIRE SPEED IP MULTICAST FORWARDING



(57) Abstract: In a switch including a plurality of ports, an IP Multicast packet arriving on an ingress port a copy of the receiver packet is forwarded to each member of the IP Multicast group at wire-speed. The packet is bridged once to a given egress port and may be routed multiple times out of the egress port. If multiple subnets exist on an egress port, each subnet that requires a copy of the packet will receive the packet with its VLAN ID included in the packet. The received IP Multicast packet for an IP Multicast group is stored in memory, a pointer to the location of the packet in memory is stored for each port to which a copy of the packet is to be forwarded. An IP Multicast forwarding entry is provided for the IP Multicast group. The forwarding entry includes a modification entry for each packet to be forwarded to the IP Multicast group. A copy of the stored packet is modified dependent on the modification entry and forwarded in the next available port cycle for the port.

**THIS PAGE BLANK (USPTO)**



*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**THIS PAGE BLANK (USPTO)**

-1-

## METHOD AND APPARATUS FOR WIRE SPEED IP MULTICAST FORWARDING

### BACKGROUND OF THE INVENTION

5 In a computer network, a networking switch receives a data packet at an ingress port connected to the switch and forwards the data packet to an egress port connected to the switch. The switch determines the egress port to which the data packet is forwarded dependent on a destination address included in the data packet received at the ingress port.

10 A destination node connected to an egress port may be a member of an IP Multicast group. A destination node joins an IP Multicast group in order to receive IP Multicast data such as a real-time video. A switch receiving an IP Multicast data packet at an ingress port must forward a copy of the received IP Multicast data packet to all members of the IP Multicast Group connected to the switch.

15 By transmitting one IP Multicast data packet to the ingress port instead of multiple unicast data packets, one for each member of the IP Multicast group, the number of data packets transmitted to the switch is reduced. However, the time to process the IP Multicast data packet in the switch is increased because the switch must determine to which egress ports the IP Multicast data packet is to be forwarded, and the number of copies to forward to each egress port.

20 The switch must forward a copy of the IP Multicast data packet to each member of the IP Multicast group. Thus, multiple copies of the IP Multicast data packet are forwarded to an egress port if multiple members of the IP Multicast group are connected to the egress port. A copy of the IP Multicast data packet is modified  
25 by the networking switch before being forwarded to a member of the IP Multicast group.

**THIS PAGE BLANK (USPTO)**

-2-

Typically, the forwarding of an IP Multicast data packet through a switch to an egress port, is performed by a host processor in the networking switch. Host processor based forwarding is slow.

#### SUMMARY OF THE INVENTION

5           A copy of an IP Multicast data packet received at an ingress port of a networking switch is forwarded at wire speed to an egress port. The IP Multicast data packet received at the ingress port is stored in memory. A copy of the stored IP Multicast data packet is forwarded to egress port forwarding logic. The egress port forwarding logic includes a modification entry for each member of the IP Multicast group associated with the IP Multicast data packet. Upon receiving a copy of an IP Multicast data packet from memory, the egress port forwarding logic modifies the copy of the IP Multicast data packet dependent on a modification entry associated with the egress port to which the IP Multicast data packet is to be forwarded.

15           Ingress port forwarding logic in the switch generates a multicast forward vector for the IP Multicast data packet. The Multicast forward vector indicates to which port the stored IP Multicast data packet is to be forwarded.

20           Port queue logic in the switch stores a pointer to the location of the IP Multicast data packet in memory on a port queue dependent on the multicast forward vector. The port queue logic removes the pointer after the IP Multicast data packet has been forwarded to all members of the IP Multicast group connected to the egress port associated with the port queue.

25           The modification entry may include an identification for a Virtual LAN, a pointer to a next modification entry, a tag indicator and a forward type entry. The pointer to a next modification entry is an index to a next modification entry for modifying the next IP Multicast data packet for the port.

          The egress port forwarding logic may include a Virtual LAN identification table and an intermediate LAN identification table. The intermediate LAN

**THIS PAGE BLANK (USPTO)**



identification table includes an index to a location in the virtual LAN table and the internal LAN table stores a Virtual LAN identification.

## BRIEF DESCRIPTION OF THE DRAWINGS

5           The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon  
10           illustrating the principles of the invention.

          Fig. 1 illustrates a network switch including IP Multicast forwarding logic for forwarding an IP Multicast packet received at an ingress port to members of an IP Multicast group connected to the switch according to the principles of the present invention;

15           Fig. 2 is a timing diagram illustrating the forwarding of IP Multicast data packets by the switch shown in Fig. 1 to members of the IP Multicast group shown in Fig. 1;

          Fig. 3A illustrates a prior art data packet which may be received on an ingress port connected to a switch;

20           Fig. 3B illustrates a prior art Ethernet Data link layer (L2) header which may be included in the data packet shown in Fig. 3A;

          Fig. 3C illustrates a prior art Internet Protocol (Network layer (L3)) header which may be included in the data packet shown in Fig. 3A;

          Fig. 3D illustrates the prior art Tag Control Information shown in Fig. 3B;

25           Fig. 4 illustrates the ingress IP Multicast forwarding logic shown in Fig. 1;

          Figs. 5 is a flowchart of the method for processing an IP Multicast data packet received on an ingress port;

          Fig. 6 illustrates the egress multicast forwarding logic shown in Fig. 1;

          Fig. 7 illustrates the Multicast Forwarding Index table shown in Fig. 6; and

30           Fig. 8 is a flowchart of the steps performed in the egress multicast forwarding logic shown in Fig. 6 for modifying the IP Multicast data packet.

**THIS PAGE BLANK (USPTO)**

## DETAILED DESCRIPTION OF THE INVENTION

Fig. 1 illustrates the members of an IP Multicast group connected to a network switch 100. The members of the IP Multicast group include destination nodes 104a-f connected to egress ports 110a-c. Source node 102 connected to an ingress port 108 is the source of IP Multicast data packets for the IP Multicast group. For example, the source node 102 may be transmitting a video to all members of the IP Multicast group. In an alternative embodiment source node 102 may be a member of the IP Multicast group. The source node 102 forwards an IP Multicast data packet 106 to the switch 100. The switch 100 forwards a copy of the IP Multicast data packet 106 to each member of the IP Multicast group, that is, destination nodes 104a-f.

The copy of the IP Multicast data packet 106 may be modified by the switch 100 before it is forwarded to each destination node 104. The copy of the received IP Multicast data packet 106 is modified dependent on a modification entry stored in the switch 100. The modification and forwarding of the copy of the received IP Multicast data packet is performed by the switch at wire-speed, that is, the modification is performed without host intervention and the egress Modified IP Multicast data packet 112 is forwarded to a destination node 104 in the next available port cycle for the egress port 110 to which the destination node 104 is connected.

As an IP Multicast data packet 106 is received on an ingress port 108, the data packet processing logic 128 extracts information from the header of the IP Multicast data packet 106 and forwards the extracted information to the ingress Multicast forwarding logic 114. The ingress Multicast forwarding logic 114 determines to which egress ports 110 the destination nodes 104 in the IP Multicast group are connected and forwards a Multicast forward vector 120 to the egress port-q logic 134 indicating the egress ports 110 to which the IP Multicast data packet is to be forwarded. The ingress Multicast forwarding logic 114 also forwards a Multicast Forwarding Index ("MFI") 132 corresponding to the IP Multicast group to

**THIS PAGE BLANK (USPTO)**

which the IP Multicast data packet 106 is to be forwarded, to the data packet processing logic 128. The MFI 132 is an index to a modification entry for each member of the IP Multicast group.

5 The data packet processing logic 128 forwards the received IP Multicast data packet 106 and the MFI 132 to memory 116. The ingress modified IP Multicast data packet 126 including the MFI 132 is stored in memory 116. The address of the ingress modified IP Multicast data packet 126 is stored in a port queue (not shown) for each egress port 110 on which the IP Multicast data packet 106 is to be forwarded. Port queues are described in co-pending U.S. Application Serial No.  
10 09/386,589, filed on August 31, 1999 entitled "Method and Apparatus for an Interleaved Non-Blocking Packet Buffer" by David A. Brown, the entire teachings of which are incorporated herein by reference.

15 The MFI 132 stored in the ingress modified IP Multicast data packet 126 is used to access a stored modification entry associated with each member of the IP Multicast group. The egress Multicast forwarding logic 118 modifies a copy of the ingress modified IP Multicast data packet 126 received from memory 116 using the contents of the stored modification entry for the member and forwards an egress modified IP Multicast data packet 112 to each member of the IP Multicast group.

20 By storing a modification entry for each member of the IP Multicast group, no host processing is required to modify and forward a copy of the IP Multicast data packet 106 to all members of the IP Multicast group. Thus, the IP Multicast data packet is forwarded at wire speed, that is, to a member of the IP Multicast group connected to a port in the next available port cycle for the port. The modification  
25 entry associated with each destination node 104 is stored in the egress Multicast forwarding logic 118 and is accessed using the MFI 132 associated with the IP Multicast Group forwarded from the Ingress Multicast forwarding logic 114. The MFI is stored in the ingress modified Multicast data packet 126 in memory 116.

**THIS PAGE BLANK (USPTO)**

As shown in Figure 1, the members of the IP Multicast group include destination nodes Da-f 104a-f. Destination nodes Da-c 104a-c are connected to egress port\_1 110a, destination nodes Dd-e 104d-e are connected to egress\_port 2 110b and destination node Df 104f is connected to egress port\_3 100c. All the destination nodes 104a-f in the embodiment shown are terminating nodes but in an alternative embodiment any one of the destination nodes 104a-f shown may be the ingress port of another switch and a copy of the IP Multicast data packet may be forwarded through the other switch to members of the IP Multicast group connected to the other switch's egress ports. Also, there may be multiple VLANs connected to the egress ports of the other switch's egress ports, with each VLAN requiring a modified copy of the IP Multicast data packet.

A data packet 106 transmitted by the source node Sa 102 is received by the network switch 100 at ingress port\_1 108. The data packet 106 includes a destination address. If the destination address is an IP Multicast address, the switch 100 forwards a modified copy of the IP Multicast data packet 112 to all members of the IP Multicast group associated with the IP Multicast address connected to the switch 100.

For the IP Multicast group shown, the switch 100 forwards six egress modified IP Multicast data packets 112a-f to egress ports 1-3 110a-c. An egress modified IP Multicast data packet 112 is forwarded to the egress port 110 in the next available port cycle for the egress port 110 after the IP Multicast data packet 106 has been received at the ingress port 102.

Three egress modified IP Multicast data packets 112a-c are forwarded to egress port\_1 110A, one to each of destination nodes Da-c 104a-c. Two egress modified IP Multicast data packet 112d-e are forwarded to egress port\_1 110b, one to each of destination nodes Dd-e 104d-e. One egress modified IP Multicast data packet 112f is forwarded to egress port\_2 110c to destination node Df 104f.

**THIS PAGE BLANK (USPTO)**



Upon identifying the received IP Multicast data packet 106, the data packet processing logic 128 forwards a portion of the header of the IP Multicast data packet 130 to ingress Multicast forwarding logic 114. The ingress Multicast forwarding logic 114 determines the egress ports\_1-3 110a-c to which a copy of the IP Multicast data packet 106 is to be forwarded and generates a Multicast forward vector 120. The Multicast forward vector 120 is forwarded to the egress port\_q logic 134. The ingress Multicast forwarding logic 114 selects a stored Multicast Forwarding Index ("MFI") 132 dependent on the portion of the header of the IP Multicast data packet 130 and forwards the MFI 132 to the data packet processing logic 128. The data packet processing logic 128 stores the MFI 132 with the received IP Multicast data packet 106 in the ingress modified IP Multicast data packet 126 and forwards the ingress modified IP Multicast data packet 126 on data-in 122 to memory 116. The ingress modified IP Multicast data packet 126 is stored in memory 116.

At the next port cycle for each of the egress ports\_1-3 110a-c, the ingress modified IP Multicast data packet 126 is forwarded from memory 116 on data-out 124 to the egress Multicast forwarding logic 118 dependent on the address 136 forwarded from egress port\_q logic 134. The egress Multicast forwarding logic 118 modifies the copy of the ingress modified data packet 126 dependent on the MFI 132 included in the ingress modified data packet 126. The egress modified data packets 112a-f are forwarded on the respective egress ports\_1-3 110a-c. After all the egress modified data packets 112a-f have been forwarded, the ingress modified data packet 126 stored in memory 116 may be overwritten.

Fig. 2 is a timing diagram illustrating wire speed forwarding of an IP Multicast data packet 106 through the switch 100 to members of the IP Multicast group shown in Fig. 1. Fig. 2 is described in conjunction with Fig. 1. The IP Multicast group shown in Fig. 1 includes six members, Da-Df 104a-f connected to the switch 100. The switch 100 includes four ports, ingress\_port 1 108 and egress ports 1-3 110a-b. However, the invention is not limited to four ports or six IP Multicast group members as shown. It is well-known to those skilled in the art that

**THIS PAGE BLANK (USPTO)**

the configuration in Fig. 1 may be expanded to include greater than four ports, and an IP Multicast group with greater than six members.

In the timing diagram, it is assumed that each IP Multicast data packet 106 received at ingress port\_1 108 is the minimum data packet size, for example, 64 bytes, for an Ethernet data packet because the memory 116 is 64 bytes wide. However, the invention is not limited to a 64 byte packet size, the number of bytes in the data packet may be greater than the width of the memory 116.

It is also assumed that the 64 byte IP Multicast data packet 106 is received at an ingress port 108 (Fig. 1) during a data segment time slot 200. The data segment time slot 200 as shown includes 4 port cycle time slots 202 one for each port connected to the switch 100 because the switch shown in Fig. 1 includes four ports connected to the switch 100, that is, one ingress port 108 and three egress ports 110a-c. The maximum number of ports supported by the switch 100 is dependent on the minimum port cycle time slot 202 and the maximum data segment time slot 200 and thus is not limited to the four ports shown in Fig. 1. In a switch 100 with greater than four ports the data segment time slot 200 would include a port cycle time slot 202 for each port connected to the switch 100. For example, for an Ethernet data packet, the minimum data packet size is 64 bytes. Thus, the maximum data segment time slot 200 is the time it takes to receive 64 bytes at an ingress port and the minimum port cycle time slot 202 is dependent on the time to access the 64 bytes from memory 116 (Fig. 1). Data segment time slots and port cycle time slots are described in co-pending U.S. Application Serial No. 09/386,589, filed on August 31, 1999 entitled "Method and Apparatus for an Interleaved Non-Blocking Packet Buffer" by David A. Brown, the entire teachings of which are incorporated herein by reference.

A copy of an IP Multicast data packet 106 arriving at ingress port\_1 108 is forwarded in the next available port cycle time slot 202 to each member of the IP Multicast group, that is, it is forwarded at wire speed. If only one member of the IP Multicast group is connected to the port the copy of the IP Multicast data packet is

**THIS PAGE BLANK (USPTO**

forwarded in the next port cycle time slot 202. If more than one member of the IP Multicast group is connected to a port, the next IP Multicast data packet arriving at the ingress port 108 is forwarded after the previous IP Multicast data packet has been forwarded to all members of the IP Multicast group connected to the port.

5           The timing diagram shows four IP Multicast data packets arriving at ingress port\_1 108 from source node-a 102 in data segment time slots 200a-d. The first IP Multicast data packet arrives in data segment time slot 200a. In the second data segment time slot 200b, the IP Multicast data packet is forwarded to egress port\_1 110a in port cycle 202b, to egress port\_2 110b in port cycle 202c and to egress  
10 port\_3 110c in port cycle 202c. As the first IP Multicast data packet is being forwarded, the second IP Multicast data packet is being received during data segment time slot 200b.

          At time 230, ingress modified IP Multicast data packet \_1 126 (Fig. 1) is stored in memory 116 (Fig. 1) and ingress multicast forwarding logic 114 (Fig. 1)  
15 has forwarded the multicast forward vector 120 (Fig. 1) for IP Multicast data packet \_1 106 (Fig. 1) to egress multicast forwarding logic 118 (Fig. 1).

          At time 206, the start of port cycle time slot 202b, a copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is forwarded from memory 116 (Fig. 1) on  
20 data-out 124 (Fig. 1) to egress Multicast forwarding logic 118 (Fig. 1) dependent on the address 136 forwarded from egress port\_q logic 134 (Fig. 1). Egress Multicast forwarding logic 118 (Fig. 1) modifies the received copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) dependent on the MFI 132 (Fig. 1). Egress modified IP Multicast data packet \_1 112A is forwarded through egress port \_1 110A  
25 to destination node Da 104a.

          At time 208, the start of port cycle time slot 202c, and at time 210, the start of port cycle time slot 202d, a copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is forwarded from memory 116 (Fig. 1) on data-out 124 (Fig. 1) dependent on the address 136 (Fig. 1) forwarded from egress port\_q logic 134 (Fig.

**THIS PAGE BLANK (USDTG)**

1). The received copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is modified by egress Multicast forwarding logic 118 dependent on the associated MFI 132 (Fig. 1) and forwarded through egress port 2 110b to destination node Dd 104d and egress port 3 110c to destination node Df 104f.

5           At time 212, a copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is modified and forwarded to destination node Db 104b (Fig. 1) and at time 218 a copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is modified and forwarded to destination node Dc 104c connected to egress port \_2 110b (Fig. 1). At time 214, a copy of ingress modified IP Multicast data packet \_1 126 (Fig. 1) is modified and forwarded to destination node Dd 104d connected to egress port \_1 110b (Fig. 1).

          At time 232, ingress modified IP Multicast data packet \_2 126 (Fig. 1) is stored in memory 116 (Fig. 1) and ingress Multicast forwarding logic 114 (Fig. 1) has forwarded the Multicast forward vector 120 (Fig. 1) for IP Multicast data packet \_2 106 (Fig. 1) to egress Multicast forwarding logic 118 (Fig. 1). At time 234, ingress modified IP Multicast data packet \_3 126 (Fig. 1) is stored in memory 116 (Fig. 1). At time 236, ingress modified IP Multicast data packet \_4 126 (Fig. 1) is stored in memory 116 (Fig. 1).

          Only one egress modified IP Multicast data packet \_1 112 (Fig. 1) is forwarded to egress port \_3 110c because only one member of the IP Multicast group, that is, destination node 104f is connected to egress port \_3 110c. Thus, at time 216, in port cycle time slot 202d, egress modified IP Multicast data packet \_2 112f is forwarded to egress port \_3 110c. At time 222, in port time cycle 202d, egress modified IP Multicast data packet \_3 112f is forwarded to egress port \_3 110c. At time 228, in port cycle time slot 3 202d, egress modified IP Multicast data packet \_4 112f is forwarded to egress port \_3 110c (Fig. 1).

          Thus, IP Multicast data packet \_1 106 (Fig. 1) received at ingress port \_0 108 (Fig. 1), is forwarded through the switch 100 (Fig. 1) at wire speed without host

**THIS PAGE BLANK (USPTO)**



-11-

processing in the next available port cycle time slot 202 to destination nodes 104a-f (Fig. 1) connected to egress ports \_1-3 110a-c (Fig. 1).

Egress modified IP Multicast data packet\_2 112 (Fig. 1) is forwarded on egress port\_2 110b to destination node De 104e and destination node Df 104f. At time 220, egress modified IP Multicast data packet\_2 112d is forwarded to destination node Dd 104d connected to egress port\_2 110b. At time 226, egress modified IP Multicast data packet\_2 112e (Fig. 1) is forwarded to destination node De 104e connected to egress port\_2 110b.

At time 224, after the last egress modified IP Multicast data packet\_1 112 (Fig. 1) has been forwarded, the location in memory storing ingress modified IP Multicast data packet \_1 126 (Fig. 1) may be used to store another data packet received at the ingress port 108 (Fig. 1). Thus, the ingress modified IP Multicast data packet 126 (Fig. 1) received at the ingress port 108 (Fig. 1) is stored in memory 116 (Fig. 1) until egress modified IP Multicast data packet 112 (Fig. 1) has been forwarded to all members of the IP Multicast group.

To determine if all egress modified IP Multicast data packets 112a-f (Fig. 1) have been forwarded, the number of port queues (not shown) that the ingress modified IP Multicast data packet 126 (Fig. 1) has been queued on associated with each ingress modified IP Multicast data packet 126 (Fig. 1) is stored in a port queue counter register (not shown). The counter register is decremented each time all egress modified IP Multicast data packets 112 (Fig. 1) associated with an egress port 110 (Fig. 1) have been forwarded to the egress port 110 (Fig. 1). For example, the port queue count is decremented after one egress modified IP Multicast data packet has been forwarded on egress port\_3 110c. The number of port queues is determined from the Multicast forward vector 120 (Fig. 1).

The prior art data packet received on ingress port 108 (Fig. 1) is described in conjunction with Figs. 3A-3D in order to describe the modifications that the forwarding logic 114, 118 (Fig. 1) and data packet processing logic 126 (Fig. 1) may

**THIS PAGE BLANK (USP)**

perform on an IP Multicast packet 106 (Fig. 1) before egress modified IP Multicast data packets 112a-f (Fig. 1) are forwarded to egress ports 110a-c.

Fig. 3A illustrates a prior art data packet which may be received on an ingress port 108 (Fig. 1) connected to the switch 100 (Fig. 1). Fig. 3B illustrates a prior art Ethernet header which may be included in the data packet shown in Fig. 3A. Fig. 3C illustrates a prior art Internet Protocol header which may be included in the data packet shown in Fig. 3A. Fig. 3D illustrates the prior art Tag Control Information field 318B in the Ethernet header shown in Fig. 3B.

Fig. 3A shows a prior art data packet 300. The data packet 300 includes a data field 310 and headers for networking layers 302, 304, 306, 308. Headers for four of the layers in the OSI model are shown, the physical layer (L1) header 302, the data link layer (L2) header 304, the networking layer (L3) header 306 and the transport layer (L4) header 308. For example, the data link layer (L2) header 304 may be Ethernet and the networking layer (L3) header 306 may be Internet Protocol ("IP"). The data packet 300 also includes a checksum 312.

Fig. 3B illustrates the format of an Ethernet data link (L2) header 304. The Ethernet data link (L2) header 304 includes a device address for the destination node 104 (Fig. 1), that is, the destination address 314, and a device address for the source node 102 (Fig. 1), that is, the source address 316, an optional Virtual Local Area Network Identification ("VLAN ID") field 318 and a length/type field 320. The VLAN ID 318 includes a Tag Protocol Identifier ("TPI") field 318A and a Tag Control Information ("TCI") field 318B. The VLAN ID field 318 provides support for VLAN switching based on IEEE 802.1Q tagging and IEEE 802.1p priority levels.

FIG. 3C illustrates the format of an IP network layer (L3) header 306. The IP network layer (L3) header 306 includes a network address for the source node 102 (Fig. 1), that is the source IP address 344, and a network address for the destination node 104 (Fig. 1), that is, the IP destination address 346. Other fields in

**THIS PAGE BLANK (USPTO)**

-13-

the IP network layer header 306 include Version 322, HLEN 324, Type of Service ("TOS") 326, Total Length 328, Identification 330, Flags 332, Fragment Offset 334, Time to Live ("TTL") 336, Protocol 340, Header Checksum 342, Options 348 and pad 350.

5           An IP Multicast data packet 106 (Fig. 1) received at the ingress port 108 (Fig. 1) connected to the switch 100 (Fig. 1) is bridged to one or more egress ports 110a-c (Fig. 1) dependent on the destination address 314 (Fig. 3B) stored in the Ethernet data link (L2) header 300 (Fig. 3A) or is routed to one or more egress ports 110 (Fig. 1) dependent on the IP destination address 346 stored the IP network layer  
10           (L3) header 306.

Fig. 3D illustrates the prior art TCI field 318b shown in Fig. 3B. The TCI field 318b includes a 3-bit user priority field 352, a 1-bit CFI field 354 and a 12-bit Virtual Local Area Network ("VLAN") identifier 356.

15           Returning to Fig. 1, the data packet processing logic 128 determines if a data packet received at ingress port\_1 108 is an IP Multicast data packet 106 by examining the network layer (L3) header 306 (Fig. 3A) included in the received data packet. The network layer (L3) header 306 (Fig. 3A) is an Internet Protocol Version 4 ("IPv4") header if the protocol field 340 (Fig. 3C) is IP, the version field 322 (Fig.  
20           3C) is IPv4 and the header checksum 342 (Fig. 3C) is valid. The IP data packet is an IP Multicast data packet 106 if the IP destination address 346 (Fig. 3C) is class D, that is, the high order four bits of the IP destination address 346 (Fig. 3C) are set to 1110.

25           If the data packet processing logic 128 determines that the data packet is an IP Multicast data packet 106, a portion of the headers of the IP Multicast data packet 130 including the IP Destination address 346 (Fig. 3C) and IP Source address 344 (Fig. 3C) in the IP networking layer (L3) header 306 (Fig. 3C), the VLAN ID 318 (Fig. 3B) in the Ethernet data link (L2) header 304 (Fig. 3B), and the ingress port

**THIS PAGE BLANK (USP 10)**

number corresponding to the ingress port 108 (Fig. 1) are forwarded to the Ingress Multicast Forwarding Logic 114.

Fig. 4 illustrates the ingress Multicast forwarding logic 114 shown in Fig. 1. The ingress Multicast forwarding logic 114 includes a Content Addressable Memory ("CAM") 400, a route mapper 402, a route table 404 and combine vector logic 406. The route table 404 includes a Multicast forwarding entry 422 corresponding to the IP Multicast Group included in the IP Multicast data packet 106 (Fig. 1) received at ingress port \_1 108 (Fig. 1). The CAM 400 and the route mapper 402 select a Multicast forwarding entry 422 dependent on the portion of the headers of the IP Multicast data packet 130 (Fig. 1) and the port number for ingress port \_1 108 (Fig. 1) forwarded from the data packet processing logic 128 (Fig. 1).

In the implementation of the Ingress Multicast forwarding logic 114 (Fig. 1) described in conjunction with Fig. 4, the CAM 400 is not sufficiently wide to store the IP Destination Address 346 (Fig. 3C), IP Source Address 344 (Fig. 3C), ingress port number and VLAN ID 318 (Fig. 3B). Thus, the route mapper 402 is implemented to increase the width of the CAM 400. However, it is well known by those skilled in the art that the CAM 400 may be expanded to include the Route Mapper 402.

The search for the Multicast Forwarding entry 422 in the Route Table 404 corresponding to the IP Multicast packet 106 is performed in two stages. First, the CAM 400 is searched for an entry dependent on the IP destination address 346 (Fig. 3C) and the IP Source Address 344 (Fig. 3C) fields in the IP network layer (L3) header 306 (Fig. 3C). If a matching entry is found in the CAM 400, a match address 410 is forwarded to the route mapper 402. The route mapper 402 forwards a Multicast Path Code ("MPC") 422 corresponding to the match address 410 to the CAM 400.

Second, the CAM 400 is searched for an entry dependent on the MPC 422, the VLAN ID 318 (Fig. 3B) and the ingress port number. If a matching entry is

**THIS PAGE BLANK (USPTO)**



-15-

found in the CAM 400, a match address 410 is forwarded to the route mapper 402. The route mapper 402 forwards an Associated Data Index ("ADI") corresponding to the match address 410 to the route table 404. An IP Multicast forwarding entry 422 is stored at a location in the route table 404 identified by the ADI 412. The IP  
5 Multicast forwarding entry 422 identifies the IP Multicast group to which the IP Multicast data packet 106 (Fig. 1) is to be forwarded by the switch 100 (Fig. 1).

The CAM 400 is searched with the IP source address 344 (Fig. 3C), the IP destination address 346 (Fig. 3C), the VLAN ID 318 (Fig. 3B) and the ingress port number 108 (Fig. 1) in order to prevent duplicating forwarding of the IP Multicast data packet. For example, if a source node is connected to more than one port, that is, destination node x is connected to port y and port x and an IP Multicast data packet is received from destination node x on port y, the data packet is forwarded to all members of the IP Multicast group. However, if the IP Multicast data packet is received from destination address x on port z it is discarded because destination x is  
10 connected to more than one port and thus the IP Multicast data packet received on port z is a duplicate of the IP Multicast data packet received on port y.

Each Multicast forwarding entry 422 in the Route Table 404 includes a bridge vector 416, a route vector 414 and a Multicast Forwarding Index 132 ("MFI"). The bridge vector 416 and route vector 414 are forwarded to combine  
20 vectors logic 406. The MFI 132 is forwarded to the data packet processing logic 128 (Fig. 1).

The combine vector logic 406 generates a Multicast forward vector 120 by performing a logical OR function on the route vector 414 and the bridge vector 416. The Multicast forward vector 120 is forwarded to the egress port\_q logic 134 (Fig.  
25 1). The Multicast forward vector 120 selects one or more egress ports to which the IP Multicast data packet 106 (Fig. 1) is to be forwarded.

The first three bytes of the destination address 314 (Fig. 3B) in the Ethernet data link (L2) header 304 (Fig. 3B) of the IP Multicast data packet 106 (Fig. 1)

**THIS PAGE BLANK (USPTO)**

-16-

received at ingress port\_1 108 are set to "0x01-00-53". Before the IP Multicast data packet 106 is forwarded on data-in 122 (Fig. 1) to memory 116 (Fig. 1), the data packet processing logic 128 (Fig. 1) writes the MFI 132 forwarded from ingress Multicast logic 114 (Fig. 1) into the second and third bytes of the destination address 314 (Fig. 3B) as follows: "0x01-0M-F1". The ingress modified IP Multicast data packet 126 (Fig. 1) is stored in memory 116 (Fig. 1). Egress Multicast forwarding logic 118 (Fig. 1) uses the MFI 132 included in the ingress modified IP Multicast data packet 126 (Fig. 1) to modify the ingress modified IP Multicast data packet 126 (Fig. 1) before forwarding an egress modified IP Multicast data packet 112 (Fig. 1) to an egress port 110 (Fig. 1).

The IP Multicast forward vector 120 determines to which egress ports 110a-c (Fig. 1) the IP Multicast data packet 106 (Fig. 1) is to be forwarded. The location of the ingress modified Multicast data packet 126 (Fig. 1) in memory 116 (Fig. 1) is stored on an egress port queue (not shown) in egress port\_q logic 134 (Fig. 1) dependent on the state of a corresponding egress port queue bit (not shown) implemented in the ingress multicast forwarding logic 114 shown in Fig. 5 in the Multicast forward vector 420.

Fig. 5 is a flowchart of the method implemented in the ingress multicast forwarding logic 114 shown in Fig. 4 for processing an IP Multicast data packet 106 (Fig. 1) received at an ingress port 108 (Fig. 1). The method is described in conjunction with Figs. 1-4.

At step 500, the ingress multicast forwarding logic 114 (Fig. 1) receives a portion of the IP Multicast data packet 130 from the data packet processing logic 128 (Fig. 1). Processing continues with step 502.

At step 502, the ingress Multicast forwarding logic 114 (Fig. 4) initializes the route vector 414 (Fig. 4) and the bridge vector 416 (Fig. 4) in the combine vectors logic 406 (Fig. 4).

**THIS PAGE BLANK (USPTO)**

-17-

At step 504, a search is performed in the CAM 400 (Fig. 4) for an entry matching the IP destination address 346 (Fig. 3C) and the IP source address 344 (Fig. 3C) forwarded to the CAM 400 (Fig. 4) from the data packet processing logic 128 (Fig. 1). If a match address 410 (Fig. 4) is found in the CAM 400 (Fig. 4),  
5 processing continues with step 506. If not, processing continues with step 518.

At step 506, the match address 410 (Fig. 4) is forwarded to the route mapper 402 (Fig. 4). The route mapper 402 (Fig. 4) forwards a Multicast Path Code ("MPC") 422 (Fig. 4) corresponding to the match address 410 (Fig. 4) to the CAM 400 (Fig. 4). Processing continues with step 508.

10 At step 508, a second search is performed in the CAM 400 (Fig. 4) for an entry dependent on the MPC 422, and the VLAN ID 318 (Fig. 3B) from the Ethernet data link (L2) header 304 (Fig. 3B) and ingress port number. If a match address 410 (Fig. 4) is found in the CAM 400 (Fig. 4) processing continues with step 510. If not, processing continues with step 518.

15 At step 510, the match address 410 (Fig. 4) is forwarded to the route mapper 402 (Fig. 4). The route mapper 402 (Fig. 4) forwards an ADI corresponding to the match address 410 (Fig. 4) to the route table 404. The IP Multicast forwarding entry 422 (Fig. 4) for the IP Multicast data packet 106 (Fig. 1) is stored at a location  
20 in the route table 404 identified by the ADI 412. Processing continues with step 512.

At step 512, the MFI 132 (Fig. 4) stored in the Multicast Forwarding entry 422 (Fig. 4) is forwarded to data packet processing logic 128 (Fig. 1) and the bridge vector 416 (Fig. 4) stored in the Multicast forwarding entry 422 (Fig. 4) is forwarded  
25 to the combine vectors logic 406 (Fig. 4). Processing continues with step 514.

At step 514, if the TTL field 336 (Fig. 3C) in the IP network layer (L3) header 306 (Fig. 3C) in the IP Multicast data packet 106 (Fig. 1) forwarded to the ingress Multicast forwarding logic 114 (Fig. 1) on the portion of the IP Multicast

**THIS PAGE BLANK (USPTO)**

-18-

data packet 130 (Fig. 1) is greater than one processing continues with step 516. If not, processing continues with step 518.

At step 516, the route vector 414 (Fig. 4) stored in the Multicast forwarding entry 422 (Fig. 4) is forwarded to the combine vectors logic 406 (Fig. 4). Processing continues with step 518.

At step 518, the ingress port number corresponding to the ingress port 108 (Fig. 1) at which the IP Multicast data packet 106 (Fig. 1) was received is removed from the bridge vector 416 (Fig. 4). Processing continues with step 519.

At step 519, the bridge vector 416 (Fig. 4) and a VLAN membership vector (not shown) are combined. Processing continues with step 520.

At step 520, the bridge vector 416 (Fig. 4) and the route vector 414 (Fig. 4) are combined by performing a logical OR function to provide a Multicast forward vector 120 (Fig. 4) for the IP Multicast group. Processing continues with step 522.

At step 522, the Multicast forward vector 422 is forwarded to the egress port\_q logic 134 (Fig. 1).

If no matching entry is found in the CAM 400 (Fig. 4) for the IP Multicast data packet 106 (Fig. 1), that is, the IP Multicast group is unknown, the Multicast data packet 106 (Fig. 1) is forwarded to the host port (not shown) so that the host port may add an entry for the IP Multicast group in the CAM 400 (Fig. 4), Route Mapper 402 (Fig. 6) and Route Table 404 (Fig. 4). The Multicast forward vector 422 (Fig. 4) is selected so that the IP Multicast packet 106 is forwarded (Fig. 1) to the host port. The Multicast forward vector 422 (Fig. 4) is also selected so that the IP Multicast data packet 106 (Fig. 1) for the unknown IP Multicast group is bridged to the VLAN on which the IP Multicast data packet 106 (Fig. 1) was received.

**THIS PAGE BLANK (USPTO)**



Fig. 6 illustrates the egress Multicast forwarding logic 118 shown in Fig. 1. The Egress Multicast forwarding logic 118 (Fig. 1) includes MFI detect logic 602, editor logic 600 and an MFI table 604. The MFI table 604 includes an MFI entry 612.

5           As a copy of an ingress modified IP Multicast data packet 126 (Fig. 1) stored in memory 116 (Fig. 1) forwarded on data-out 124 (Fig. 1) is received by the egress IP Multicast processing logic 118, the MFI detect logic 602 detects the MFI 132 included in the ingress modified IP Multicast data packet 126 (Fig. 1) and forwards the MFI 132 to the MFI Table 604. The MFI 132 is an index to the MFI entry 612  
10           in the MFI table 604 corresponding to the IP Multicast group. The MFI entry 612 determines to which egress ports 112 (Fig. 1) the first copy of the IP Multicast data packet 106 (Fig. 1) is to be forwarded. If a subsequent IP Multicast data packet 106 is to be forwarded on an egress port 110, a next MFI 132' is included in the MFI entry 612 in the MFI table 604. The next MFI 132' is forwarded to the MFI detect  
15           logic 602 to select the next MFI entry 612. For example, a next MFI 132' is stored with the MFI entry 612 for egress modified IP Multicast data packet 112a (Fig. 1) to select the next MFI entry 612' for egress modified IP Multicast data packet 112b (Fig. 1).

          The ingress modified IP Multicast data packet 126 (Fig. 1) is modified by  
20           the editor logic 600 dependent on Multicast parameters 610 included in the MFI entry 612. Multicast parameters 610 stored in the MFI entry 612 are forwarded to the editor logic 600. Thus, each ingress modified IP Multicast data packet 126 (Fig. 1) is edited dependent on the destination node 104a-f (Fig. 1) to which it is being forwarded. By providing an MFI entry 612 in the MFI table 604, the ingress  
25           modified IP Multicast data packet 126 (Fig. 1) may be edited to provide an egress modified IP Multicast data packet 112 (Fig. 1) and forwarded at wire speed to each member of the IP Multicast group connected to an egress port 110a-c (Fig. 1).

Fig. 7 illustrates the MFI table 604 shown in Fig. 6 for a switch 100 with four ports as shown in Fig. 1. The MFI table 604 (Fig. 6) includes an IVID table

**THIS PAGE BLANK (USPTO)**

-20-

700 and a VLAN ID table 702. An MFI entry 612' includes an IVID entry 718 in the IVID table 700 and a VLAN ID entry 710 in the VLAN ID table 702. An IVID entry 718 in the IVID table 700 includes a port field 714a-d for each port connected to the switch 100 (Fig. 1). The switch 100 as shown in Fig. 1 includes four ports:  
5 ingress port \_1 108 and egress ports \_1-3 110a-c.

Each port field 714 includes a Bridge/Route field 708, a Next field 706 and an IVID field 704. The IVID entry 718 also includes a next MFI field 716. The next MFI field stores a next MFI 132' if any next field 706a-d in the IVID entry 718  
10 indicates that another egress modified IP Multicast data packet 112 (Fig. 1) is to be forwarded to a port connected to the switch 100. (Fig. 1) The IVID entry 718 as shown is for a switch 100 (Fig. 1) with four ports, as shown in Fig. 1. The number of port fields 714 is not limited to the four shown, a port field 714 is provided for each port connected to the switch 100 (Fig. 1). For example, for a switch with  
15 sixty-four ports, the IVID entry 718 includes sixty four port fields 714.

The IVID field 704 stores an index to a VLAN ID entry 716 in the VLAN ID table 702 corresponding to the IVID field 704 if there is a valid VLAN ID entry 716 for the port. The VLAN ID entry 716 in the VLAN ID table 702 includes a VLAN ID field 710 and a Tag field 712.

20 In an alternate embodiment the IVID entry 718 in the IVID Table 700 may include the VLAN ID field 710 and tag field 712 stored in the VLAN ID entry 716, in which case the IVID field 704 would not be required. However, storing the IVID entry 718 and the VLAN ID entry 710 separately, that is, in the IVID table 700 and the VLAN ID table 702 reduces the total memory required to implement the MFI  
25 table 604 (Fig. 6) because every ingress modified IP Multicast data packet 126 may not have a corresponding VLAN ID entry 716.

The Bridge/Route field 708 is implemented as a single bit. The state of the Bridge/Route field 708 determines whether the ingress modified IP Multicast data packet 126 (Fig. 1) is to be bridged or routed to a port. The Next field 706 is

**THIS PAGE BLANK (USPTO)**

implemented as a single bit. The state of the next field 706 determines whether another egress modified IP Multicast data packet 112 (Fig. 1) is to be forwarded on the port. The IVID field 704 provides an index to the VLAN ID table 702. The number of bits in the IVID field 704 is dependent on the size of the VLAN ID table 702. For example, for a VLAN ID Table 706 with 256 VLAN ID entries 716 the VLAN ID index 714 includes eight bits.

The VLAN ID 710 contains a valid VLAN ID 710 if the ingress modified IP Multicast Data packet 126 (Fig. 1) is to be routed to the port. If the Tag field 712 indicates that the egress modified IP Multicast data packet 112 (Fig. 1) is not tagged, the editor logic 600 (Fig. 6) deletes the tag 318b (Fig. 3B) from the VLAN ID 318 (Fig. 3B) in the Ethernet data link layer (L2) header 304 (Fig. 3B).

Having determined that a VLAN ID 710 is required and obtaining the VLAN ID 710 from the VLAN ID table 702, the Multicast parameters 610 (Fig. 6) are forwarded to the editor logic 600 (Fig. 6). The editor logic 600 (Fig. 6) uses the Multicast parameters 610 to modify the ingress modified IP Multicast data packet 126 (Fig. 6), and after modifying the editor logic 600 (Fig. 6) forwards the egress modified IP Multicast data packet 112 to the egress port 110 (Fig. 1).

If the Bridged/Routed field 708 in the port field 714 is set to bridge. The editor logic 600 (Fig. 6) adds or removes the VLAN ID 318 (Fig. 3C) and restores the destination address 314 (Fig. 3B) in the Ethernet data link layer (L2) header 304 (Fig. 3B) to its original value.

If the Bridged/Routed field 708 in the port field 714 is set to routed. The editor logic 600 (Fig. 6) overwrites the router's source address in the source address 314 (Fig. 3B) in the Ethernet data link layer (L2) header 304 (Fig. 3B), decrements the TTL 336 (Fig. 3C) and generates an IP check sum and stores it in header checksum 342 (Fig. 3C). The editor logic 600 (Fig. 6) also restores the IP destination address 346 (Fig. 3C) to the original value and adds the TCI 318b (Figure 3B) tag.

**THIS PAGE BLANK (USPTO)**

The egress Multicast forwarding logic 118 (Fig. 1) forwards the IP destination address 346 (Fig. 3C) the egress modified IP Multicast data packet 112a-f (Fig. 1) to the egress ports 110a-c (Fig. 1).

5 While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

**THIS PAGE BLANK (USPTO)**



## CLAIMS

What is claimed is:

1. A switch comprising:

5 a memory storing a multicast data packet received from a port connected to the switch;

egress port forwarding logic, the egress port forwarding logic comprising:

a modification entry for each member of the multicast group associated with the multicast data packet; and

10 editing logic which modifies a copy of the multicast data packet received from the memory dependent on the modification entry.

2. The switch as claimed in Claim 1 further comprising:

15 ingress port forwarding logic which generates a multicast forward vector for the multicast data packet, the multicast forward vector indicating to which ports the stored multicast data packet is to be forwarded.

3. The switch as claimed in Claim 2 further comprising:

20 port queue logic which stores a pointer to the multicast data packet stored in the memory in at least one port queue dependent on the multicast forward vector.

4. A switch as claimed in Claim 3 wherein the port queue logic removes the pointer to the multicast data packet stored in the port queue after the stored multicast data packet has been forwarded to all members of the IP Multicast group connected to the port associated with the port queue.

25 5. A switch as claimed in Claim 1 wherein the modification entry comprises:

a virtual LAN identification;

a pointer to a next modification entry;

a tag indicator; and

a forward type indicator.

**THIS PAGE BLANK (USPTO)**

-24-

6. A switch as claimed in Claim 5 wherein the pointer to next modification entry is an index to a next modification entry for modifying a next copy of the multicast data packet.

7. A switch as claimed in Claim 1 wherein the modification entry comprises:

5 a virtual LAN entry stored in a virtual LAN identification table; and  
an intermediate entry stored in an intermediate entry stored in an intermediate LAN identification table wherein the intermediate entry includes an index to the virtual LAN entry.

8. A method for forwarding multicast data packets in a switch comprising the steps of:

10 storing in a memory, a multicast data packet received from an ingress port;  
providing egress port forwarding logic, the egress port forwarding logic including a modification entry for each member of the multicast group associated with the multicast data packet and editing logic for modifying the multicast data  
15 packet;

modifying in the egress port forwarding logic a copy of the multicast data packet received from the memory dependent on the modification entry; and  
forwarding by the egress port forwarding logic the modified copy of the multicast data packet to an egress port in a next available egress port cycle time.

20

9. The method for forwarding multicast packets in a switch as claimed in Claim 8 further comprising the step of:

selecting a multicast forward vector for the multicast data packet dependent on parameters included in the multicast data packet, the multicast forward vector  
25 indicating to which egress ports the multicast data packet is to be forwarded.

10. The method for forwarding multicast packets in a switch as claimed in Claim 9 further comprising the step of:

**THIS PAGE BLANK (USPTO)**

storing, by the packet manager a pointer to the multicast data packet stored in the memory in a port queue associated with the egress port dependent on the multicast forward vector.

5        11.     The method for forwarding multicast packets in a switch as claimed in Claim 10 further comprising the step of:

         removing, by the packet manager the pointer to the multicast data packet stored in the port queue after all modified copies of the multicast data packet are transmitted to the egress port associated with the port queue.

10

12.     The method for forwarding multicast packets in a switch as claimed in Claim 8 wherein the step of providing a modification entry comprises the step of:  
         providing a virtual LAN identification, a pointer to next entry, a tag indicator and a forward type indicator.

15        13.     The method for forwarding multicast packets in a switch as claimed in Claim 12 wherein the pointer to next entry provides an index to a next modification entry storing the parameters for a second modified copy of the multicast data packet to be sent to the egress port.

20        14.     The method for forwarding multicast packets in a switch as claimed in Claim 8 wherein the step of providing egress port forwarding logic further comprises the steps of:

         providing a virtual LAN identification table; and

         providing an intermediate LAN identification table for storing the multicast entry wherein the multicast entry includes an index to a location in the virtual LAN identification table including a virtual LAN identification.

25

15.     A switch comprising:  
         a memory storing a multicast data packet received from an ingress port connected to the switch;  
         egress port forwarding logic, the egress port forwarding logic comprising:

**THIS PAGE BLANK (USPTO)**

-26-

a modification entry for each member of the multicast group associated with the multicast data packet; and

means for modifying a copy of the multicast data packet received from the memory dependent on the modification entry.

5        16.     The switch as claimed in Claim 15 further comprising:  
             means for generating a multicast forward vector for the multicast data packet,  
             the multicast forward vector indicating to which ports the stored multicast data  
             packet is to be forwarded.

10       17.     The switch as claimed in Claim 16 further comprising:  
             means for storing a pointer to the multicast data packet stored in the memory  
             in at least one port queue dependent on the multicast forward vector.

15       18.     A switch as claimed in Claim 17 wherein the means for storing removes the  
             pointer to the multicast data packet stored in the port queue after the multicast data  
             packet has been transmitted to each member of the multicast group connected to the  
             egress port associated with the port queue.

20       19.     A switch as claimed in Claim 15 wherein the modification entry comprises:  
             a virtual LAN identification;  
             a pointer to next entry;  
             a tag indicator; and  
             a forward type indicator.

25       20.     A switch as claimed in Claim 19 wherein the pointer to next entry provides  
             an index to a next modification entry storing the parameters for a next modified  
             copy of the multicast data packet to be forwarded to the egress port.

21.     The switch as claimed in Claim 15 wherein the modification entry  
             comprises:

**THIS PAGE BLANK (USPTO)**



a virtual LAN entry stored in a virtual LAN identification table; and  
an intermediate entry stored in an intermediate entry stored in an  
intermediate LAN identification table wherein the intermediate entry includes an  
index to the virtual LAN entry.

**THIS PAGE BLANK (USPTO)**

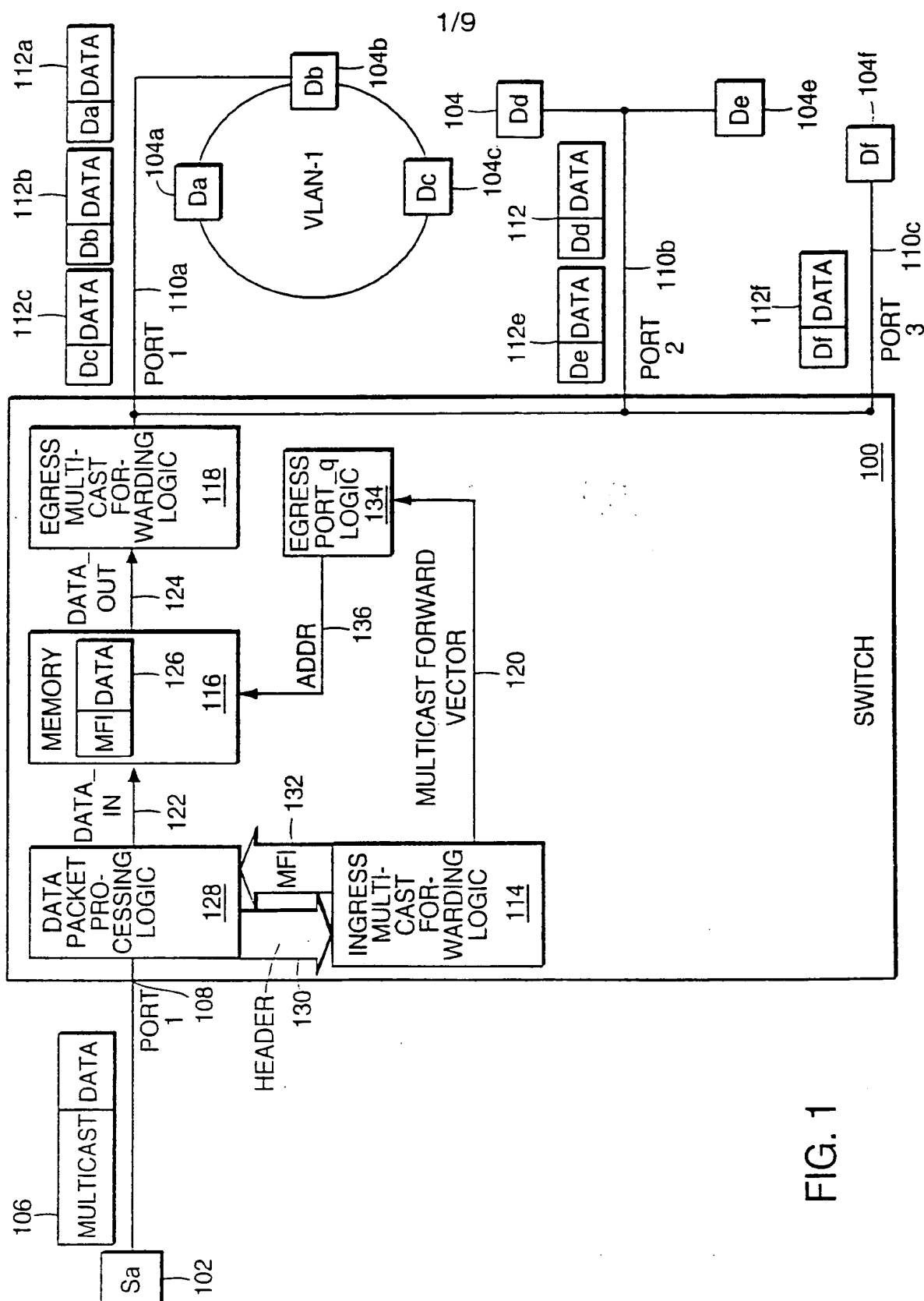


FIG. 1

**THIS PAGE BLANK (USPTO)**

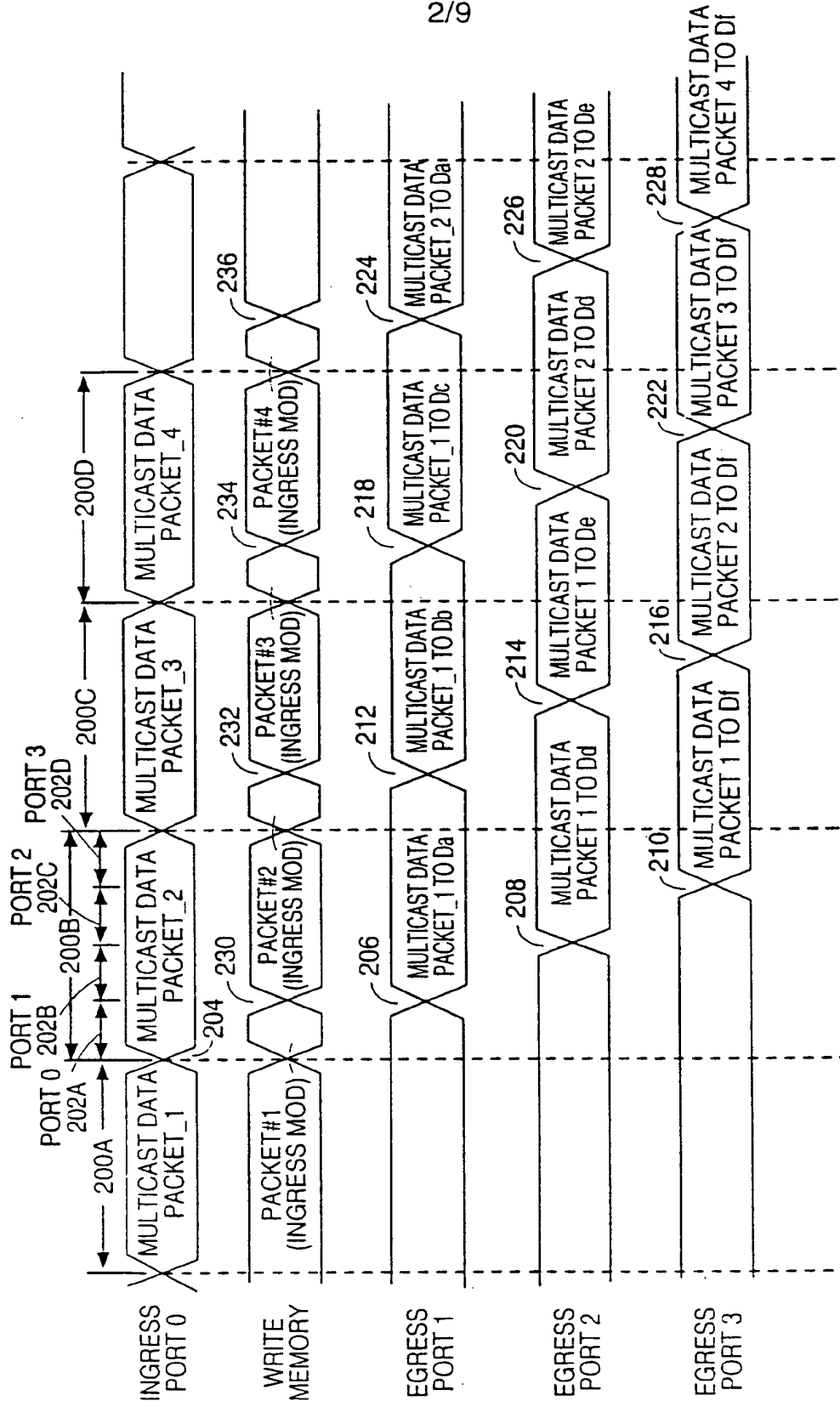
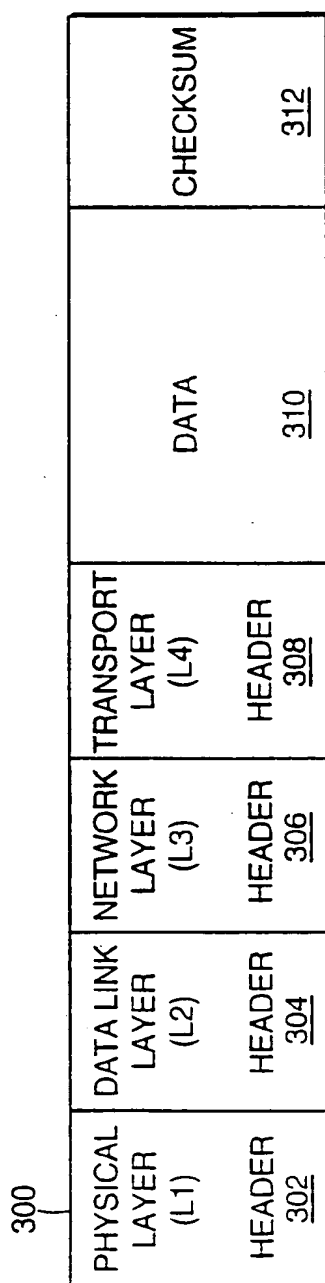
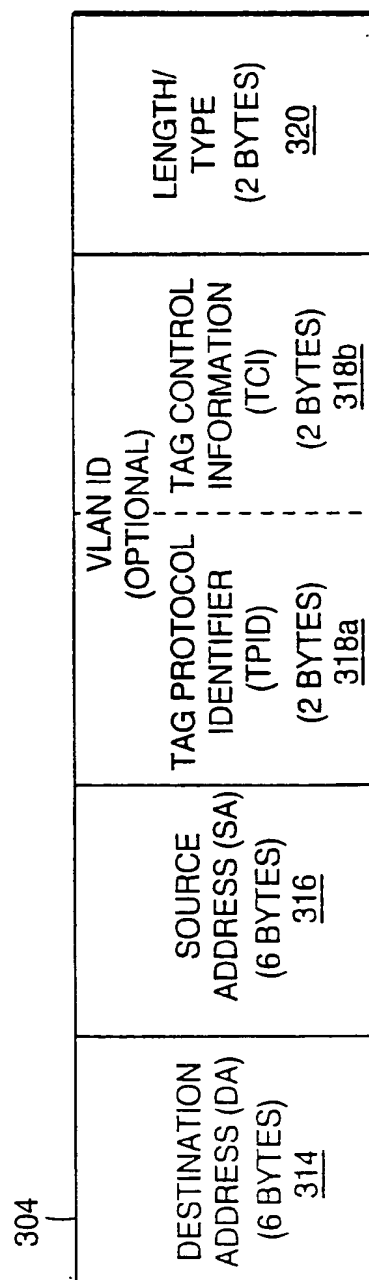


FIG. 2

**THIS PAGE BLANK (18870)**



PRIOR ART  
FIG. 3A



PRIOR ART  
FIG. 3B

**THIS PAGE BLANK (USPTO)**



4/9

306

<u>322</u> VERS	<u>324</u> HLEN	TOS <u>326</u>	<u>328</u> TOTAL LENGTH
<u>330</u> IDENTIFICATION		<u>332</u> FLAGS	FRAGMENT OFFSET <u>334</u>
<u>336</u> TTL	<u>340</u> PROTOCOL	<u>342</u> HEADER CHECKSUM	
SRC IP ADDRESS			<u>344</u>
DEST IP ADDRESS			<u>346</u>
<u>348</u> OPTIONS			<u>350</u> PAD

PRIOR ART  
FIG. 3C

318B

USER PRIORITY (3 BITS) <u>352</u>	<u>354</u> CFI (1 BIT)	<u>356</u> VLAN IDENTIFIER (12 BITS)
--	---------------------------------	--

PRIOR ART  
FIG. 3D

**THIS PAGE BLANK (USPTO)**

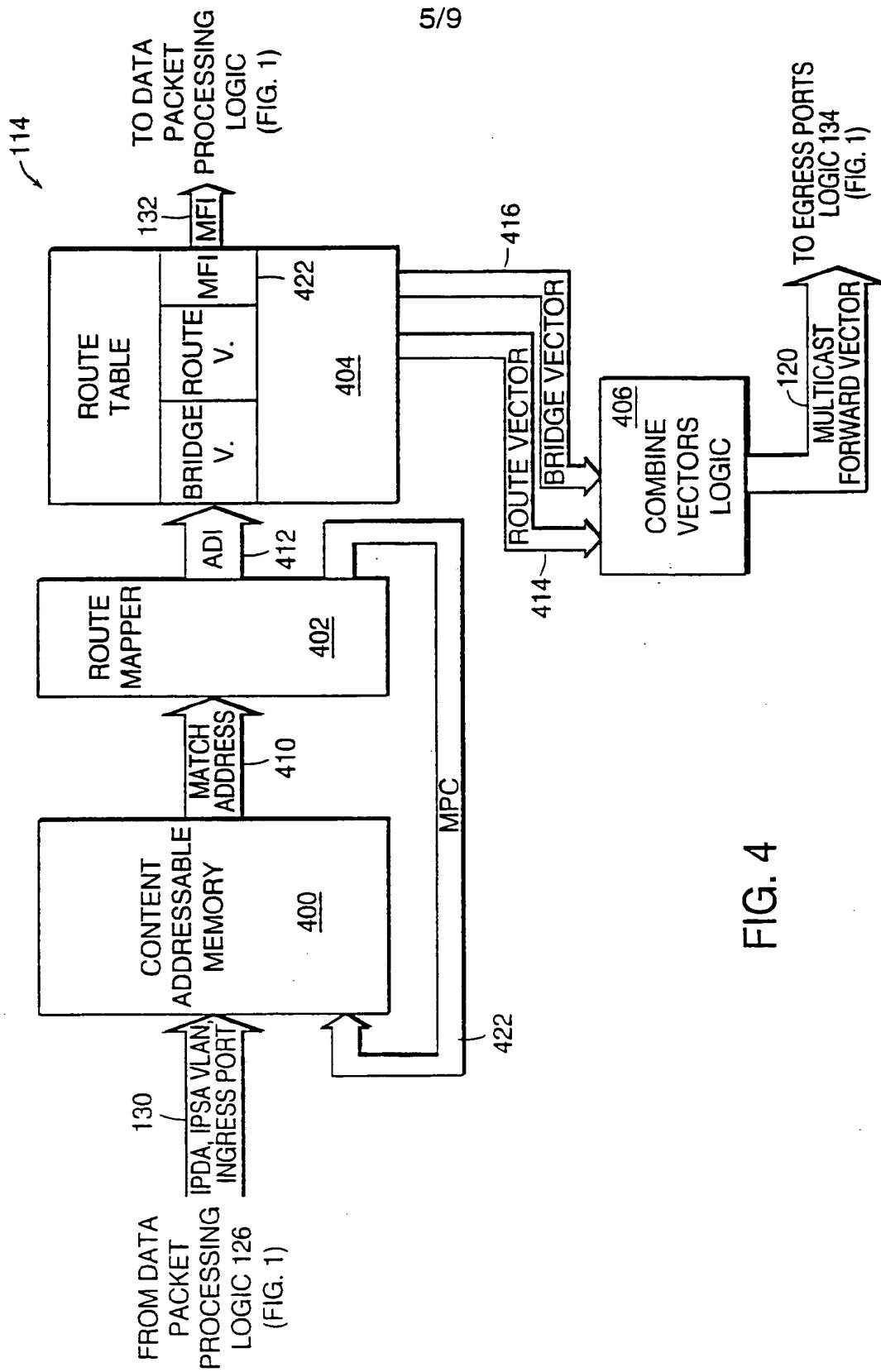


FIG. 4

**THIS PAGE BLANK (USPTO)**

6/9

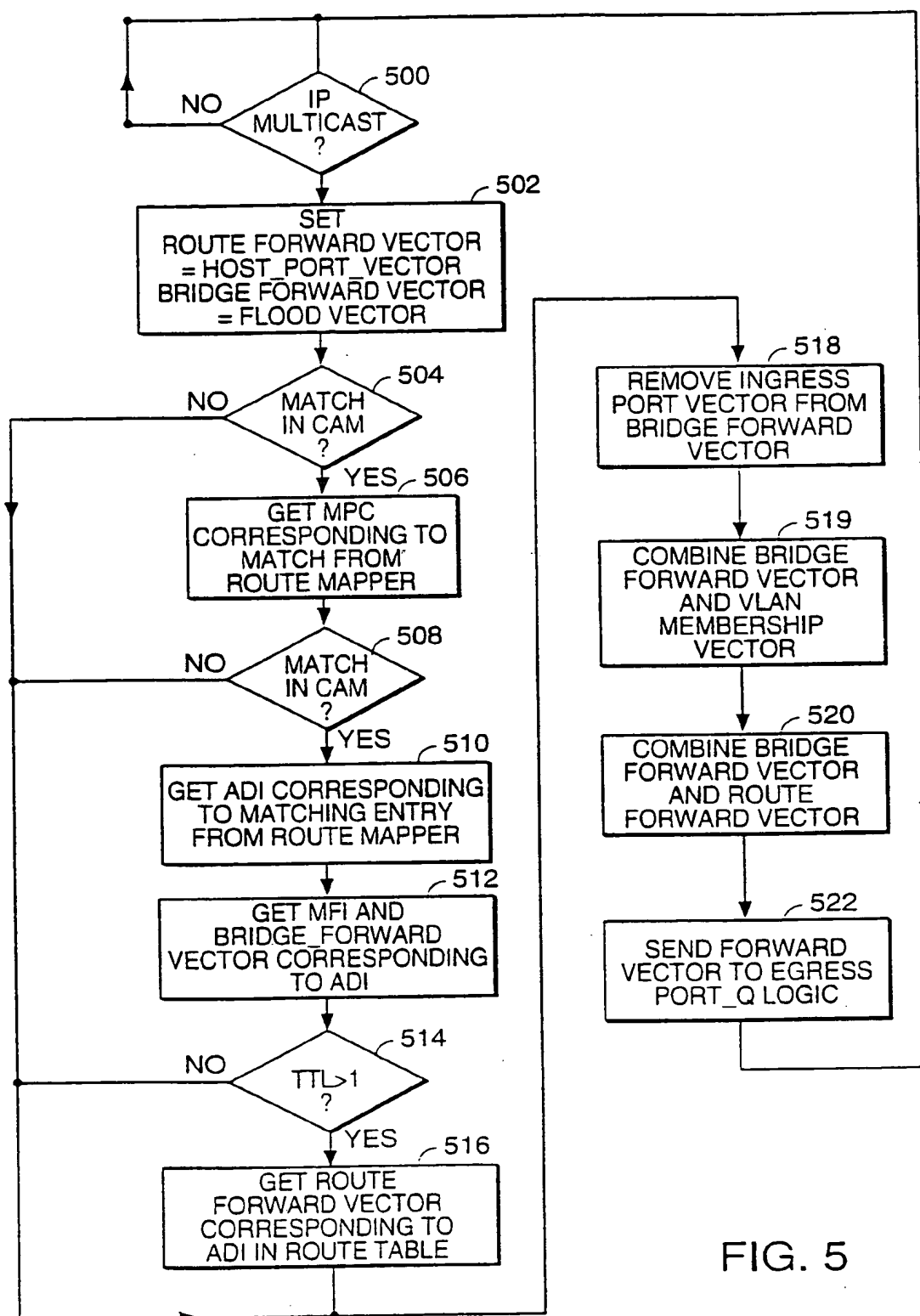


FIG. 5

**THIS PAGE BLANK (USPTO)**

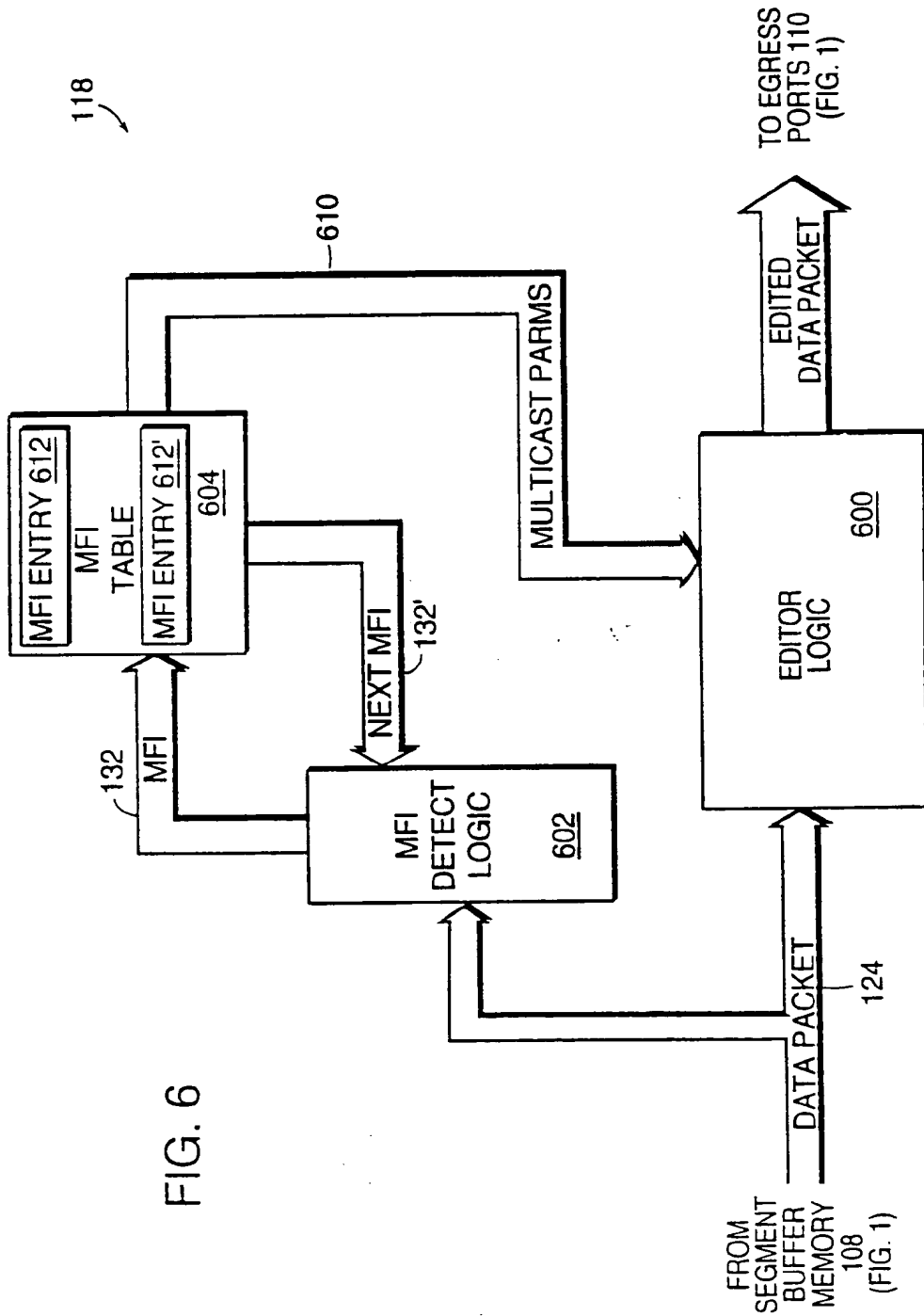
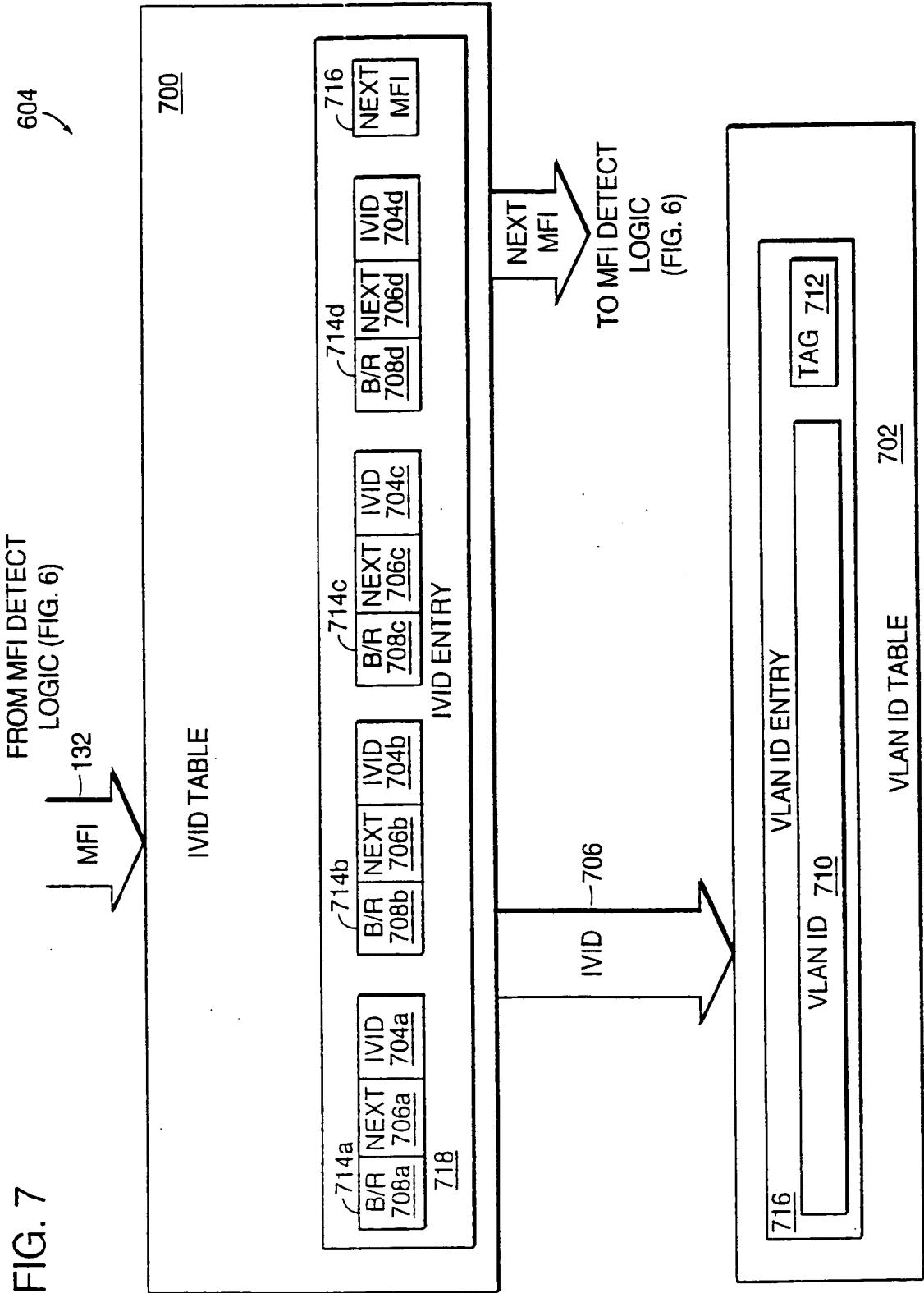


FIG. 6

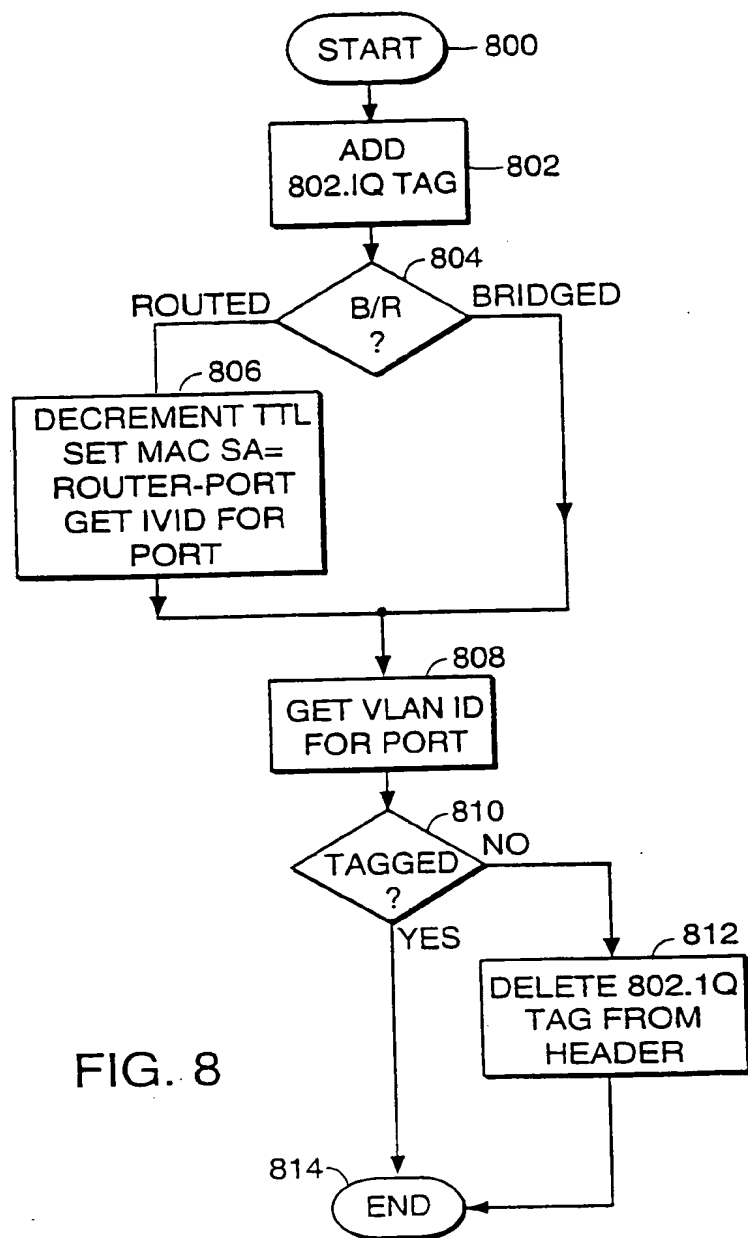
**THIS PAGE BLANK (USPTO)**





**THIS PAGE BLANK (USPTO)**

9/9





(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表 2003-516029

(P 2003-516029A)

(43) 公表日 平成15年5月7日 (2003. 5. 7)

(51) Int. Cl.<sup>7</sup> 識別記号 F I テーマコード\* (参考)  
H 0 4 L 12/56 2 6 0 H 0 4 L 12/56 2 6 0 A 5K030

審査請求 未請求 予備審査請求 有

(全 4 7 頁)

(21) 出願番号 特願2001-541181 (P2001-541181)  
(86) (22) 出願日 平成12年11月30日 (2000. 11. 30)  
(85) 翻訳文提出日 平成14年5月31日 (2002. 5. 31)  
(86) 国際出願番号 PCT/CA00/01420  
(87) 国際公開番号 WO01/041364  
(87) 国際公開日 平成13年6月7日 (2001. 6. 7)  
(31) 優先権主張番号 09/453, 344  
(32) 優先日 平成11年12月1日 (1999. 12. 1)  
(33) 優先権主張国 米国 (U S)

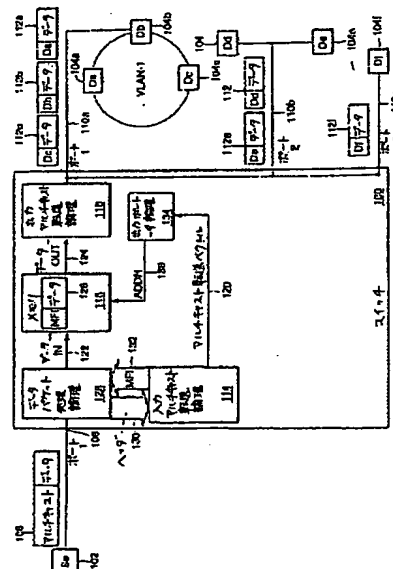
(71) 出願人 モサイド・テクノロジーズ・インコーポレ  
イテッド  
カナダ、ケイ・2・ケイ 2・エックス・1  
オンタリオ州、カナタ、ハインズ・ロー  
ド、11  
(72) 発明者 ブラウン、デイビッド・エイ  
カナダ、ケイ・0・エイ 1・エル・0 オ  
ンタリオ州、カープ、アリシア・クレッセ  
ント、110  
(74) 代理人 弁理士 深見 久郎 (外5名)

最終頁に続く

(54) 【発明の名称】 ワイヤ速度の I P マルチキャスト転送のための方法および装置

(57) 【要約】

複数個のポートを含むスイッチにおいて、入力ポートに到着する I P マルチキャストパケット、受信側パケットのコピーが、ワイヤ速度で I P マルチキャストグループの各メンバに転送される。パケットは、一旦所与の出力ポートにブリッジされ、出力ポートから複数回ルーチングされてもよい。複数のサブネットが出力ポートに存在する場合、パケットのコピーを要求する各サブネットは、パケットに含まれるその V L A N I D とともにパケットを受信する。I P マルチキャストグループのための受信された I P マルチキャストパケットは、メモリに記憶され、メモリにおけるパケットの場所へのポインタは、パケットのコピーが転送されるべき各ポートごとに記憶される。I P マルチキャスト転送エントリが、I P マルチキャストグループのために設けられる。転送エントリは、I P マルチキャストグループに転送されるべき各パケットのための修正エントリを含む。記憶されたパケットのコピーは、修正エントリに依存して修正され、ポートのための次の使用可能ポートサイクルにおいて転送される。



**【特許請求の範囲】**

【請求項1】 スイッチに接続されたポートから受信されるマルチキャストデータパケットを記憶するメモリと、

出力ポート転送論理とを含み、出力ポート転送論理は、

マルチキャストデータパケットと関連付けられるマルチキャストグループの各メンバのための修正エン트리と、

修正エントリに依存してメモリから受信されるマルチキャストデータパケットのコピーを修正する編集論理とを含む、スイッチ。

【請求項2】 マルチキャストデータパケットのためのマルチキャスト転送ベクトルを生成する入力ポート転送論理をさらに含み、マルチキャスト転送ベクトルは、記憶されたマルチキャストデータパケットがどのポートに転送されるべきかを示す、請求項1に記載のスイッチ。

【請求項3】 マルチキャスト転送ベクトルに依存して少なくとも1つのポートキューにメモリに記憶されたマルチキャストデータパケットへのポインタを記憶するポートキュー論理をさらに含む、請求項2に記載のスイッチ。

【請求項4】 記憶されたマルチキャストデータパケットが、ポートキューと関連付けられたポートに接続されたIPマルチキャストグループのすべてのメンバに転送されてしまった後、ポートキュー論理は、ポートキューに記憶されたマルチキャストデータパケットへのポインタを除去する、請求項3に記載のスイッチ。

【請求項5】 修正エントリは、

仮想LAN識別子と、

次の修正エントリへのポインタと、

タグインジケータと、

転送タイプインジケータとを含む、請求項1に記載のスイッチ。

【請求項6】 次の修正エントリへのポインタは、マルチキャストデータパケットの次のコピーを修正するための次の修正エントリへのインデックスである、請求項5に記載のスイッチ。

【請求項7】 修正エントリは、

仮想LAN識別テーブルに記憶された仮想LANエン트리と、

中間LAN識別テーブルに記憶された中間エントリに記憶された中間エントリとを含み、中間エントリは、仮想LANエントリへのインデックスを含む、請求項1に記載のスイッチ。

【請求項8】 入力ポートから受信されるマルチキャストデータパケットをメモリに記憶するステップと、

出力ポート転送論理を設けるステップとを含み、出力ポート転送論理は、マルチキャストデータパケットと関連付けられるマルチキャストグループの各メンバーのための修正エントリと、マルチデータパケットを修正するための編集論理とを含み、さらに、

出力ポート転送論理において修正エントリに依存してメモリから受信されたマルチキャストデータパケットのコピーを修正するステップと、

出力ポート転送論理により、マルチキャストデータパケットの修正されたコピーを出力ポートへ、次の使用可能出力ポートサイクル時間において転送するステップとを含む、スイッチにおいてマルチキャストデータパケットを転送するための方法。

【請求項9】 マルチキャストデータパケットに含まれるパラメータに依存してマルチキャストデータパケットのためのマルチキャスト転送ベクトルを選択するステップをさらに含み、マルチキャスト転送ベクトルは、マルチキャストデータパケットがどの出力ポートに転送されるべきかを示す、請求項8に記載のスイッチにおいてマルチキャストパケットを転送するための方法。

【請求項10】 パケットマネージャにより、マルチキャスト転送ベクトルに依存して出力ポートと関連付けられるポートキューにメモリに記憶されたマルチキャストデータパケットへのポインタを記憶するステップをさらに含む、請求項9に記載のスイッチにおいてマルチキャストパケットを転送するための方法。

【請求項11】 マルチキャストデータパケットのすべての修正されたコピーがポートキューと関連付けられる出力ポートに伝送された後、パケットマネージャにより、ポートキューに記憶されたマルチキャストデータパケットへのポインタを除去するステップをさらに含む、請求項10に記載のスイッチにおいてマ

ルキャストパケットを転送するための方法。

【請求項12】 修正エントリを設けるステップは、

仮想LAN識別子、次のエントリへのポインタ、タグインジケータ、および転送タイプインジケータを設けるステップを含む、請求項8に記載のスイッチにおいてマルチキャストパケットを転送するための方法。

【請求項13】 次のエントリへのポインタは、出力ポートに送信されるべきマルチキャストデータパケットの第2の修正されたコピーのためのパラメータを記憶する次の修正エントリへのインデックスを与える、請求項12に記載のスイッチにおいてマルチキャストパケットを転送するための方法。

【請求項14】 出力ポート転送論理を設けるステップは、

仮想LAN識別テーブルを設けるステップと、

マルチキャストエントリを記憶するための中間LAN識別テーブルを設けるステップとをさらに含み、マルチキャストエントリは、仮想LAN識別子を含む仮想LAN識別テーブル中の場所へのインデックスを含む、請求項8に記載のスイッチにおいてマルチキャストパケットを転送するための方法。

【請求項15】 スwitchに接続された入力ポートから受信されるマルチキャストデータパケットを記憶するメモリと、

出力ポート転送論理とを含み、出力ポート転送論理は、

マルチキャストデータパケットと関連付けられるマルチキャストグループの各メンバのための修正エントリと、

修正エントリに依存してメモリから受信されたマルチキャストデータパケットのコピーを修正するための手段とを含む、スイッチ。

【請求項16】 マルチキャストデータパケットのためのマルチキャスト転送ベクトルを生成するための手段をさらに含み、マルチキャスト転送ベクトルは、記憶されたマルチキャストデータパケットがどのポートに転送されるべきかを示す、請求項15に記載のスイッチ。

【請求項17】 マルチキャスト転送ベクトルに依存して少なくとも1つのポートキューにメモリに記憶されたマルチキャストデータパケットへのポインタを記憶するための手段をさらに含み、請求項16に記載のスイッチ。



【請求項18】 前記記憶するための手段は、マルチキャストデータパケットがポートキューと関連付けられる出力ポートに接続されたマルチキャストグループの各メンバに伝送されてしまった後、ポートキューに記憶されたマルチキャストデータパケットへのポインタを除去する、請求項17に記載のスイッチ。

【請求項19】 修正エントリは、  
仮想LAN識別子と、  
次のエントリへのポインタと、  
タグインジケータと、  
転送タイプインジケータとを含む、請求項15に記載のスイッチ。

【請求項20】 次のエントリへのポインタは、出力ポートに転送されるべきマルチキャストデータパケットの次の修正されたコピーのためのパラメータを記憶する次の修正エントリへのインデックスを与える、請求項19に記載のスイッチ。

【請求項21】 修正エントリは、  
仮想LAN識別テーブルに記憶された仮想LANエントリと、  
中間LAN識別テーブルに記憶された中間エントリに記憶された中間エントリとを含み、中間エントリは、仮想LANエントリへのインデックスを含む、請求項15に記載のスイッチ。

**【発明の詳細な説明】****【0001】****【発明の背景】**

コンピュータネットワークにおいては、ネットワーキングスイッチが、スイッチに接続された入力ポートでデータパケットを受信し、スイッチに接続された出力ポートにそのデータパケットを転送する。スイッチは、入力ポートで受信されたデータパケットに含まれる宛先アドレスに依存して、データパケットが転送される出力ポートを判断する。

**【0002】**

出力ポートに接続された送信先ノードが、IPマルチキャストグループのメンバーであり得る。送信先ノードは、リアルタイムビデオなどのIPマルチキャストデータを受信するためにIPマルチキャストグループを結合する。入力ポートでIPマルチキャストデータパケットを受信するスイッチは、受信されたIPマルチキャストデータパケットのコピーをスイッチに接続されたIPマルチキャストグループのすべてのメンバーに転送しなければならない。

**【0003】**

IPマルチキャストグループの各メンバーごとに1つの、複数のユニキャストデータパケットの代わりに、1つのIPマルチキャストデータパケットを入力ポートに伝送することにより、スイッチに伝送されるデータパケットの数は低減される。しかしながら、スイッチは、IPマルチキャストデータパケットがどの出力ポートに転送されるか、および、各出力ポートに転送すべきコピーの数を判断しなければならないので、スイッチにおいてIPマルチキャストデータパケットを処理する時間が増大する。

**【0004】**

スイッチは、IPマルチキャストデータパケットのコピーをIPマルチキャストグループの各メンバーに転送しなければならない。したがって、IPマルチキャストグループの複数のメンバーが出力ポートに接続されるならば、IPマルチキャストデータパケットの複数のコピーが出力ポートに転送される。IPマルチキャストデータパケットのコピーは、IPマルチキャストグループのメンバーに転送さ

れる前に、ネットワーキングスイッチによって修正される。

#### 【0005】

典型的には、スイッチを介する出力ポートへのIPマルチキャストデータパケットの転送は、ネットワーキングスイッチにおけるホストプロセッサによって行なわれる。ホストプロセッサに基づいた転送は遅い。

#### 【0006】

##### 【発明の概要】

ネットワーキングスイッチの入力ポートで受信されたIPマルチキャストデータパケットのコピーがワイヤ速度で出力ポートに転送される。入力ポートで受信されたIPマルチキャストデータパケットは、メモリに記憶される。記憶されたIPマルチキャストデータパケットのコピーが、出力ポート転送論理に転送される。出力ポート転送論理は、IPマルチキャストデータパケットと関連付けられるIPマルチキャストグループの各メンバごとに修正エントリを含む。メモリからIPマルチキャストデータパケットのコピーを受信すると、出力ポート転送論理は、IPマルチキャストデータパケットが転送されるべき出力ポートと関連付けられる修正エントリに依存してIPマルチキャストデータパケットのコピーを修正する。

#### 【0007】

スイッチにおける入力ポート転送論理が、IPマルチキャストデータパケットのためのマルチキャスト転送ベクトルを生成する。マルチキャスト転送ベクトルは、記憶されたIPマルチキャストデータパケットがどのポートに転送されるべきかを示す。

#### 【0008】

スイッチにおけるポートキュー論理が、マルチキャスト転送ベクトルに依存してポートキューのメモリにIPマルチキャストデータパケットの場所へのポインタを記憶する。IPマルチキャストデータパケットが、ポートキューと関連付けられる出力ポートに接続されたIPマルチキャストグループのすべてのメンバに転送されてしまった後、ポートキュー論理はポインタを除去する。

#### 【0009】

修正エントリは、仮想LAN (Virtual LAN) のための識別子と、次の修正エントリへのポインタと、タグインジケータと、転送タイプエントリとを含み得る。次の修正エントリへのポインタは、ポートのための次のIPマルチキャストデータパケットを修正するための次の修正エントリへのインデックスである。

#### 【0010】

出力ポート転送論理は、仮想LAN識別テーブルと、中間LAN識別テーブルとを含み得る。中間LAN識別テーブルは、仮想LANテーブル中の場所へのインデックスを含み、内部LANテーブルは、仮想LAN識別子を記憶する。

#### 【0011】

この発明の前に述べたおよび他の目的、特徴および利点は、添付の図面に記載される、この発明の好ましい実施例の以下のより特定のな説明から明らかとなり、異なった図面を通じて同様の参照符号は同様の部品を参照する。図は、必ずしも一定の比例で描かれているとは限らず、代わりにこの発明の原理を説明することに強調が置かれている。

#### 【0012】

##### 【詳細な説明】

図1は、ネットワークスイッチ100に接続されるIPマルチキャストグループのメンバを示す。IPマルチキャストグループのメンバは、出力ポート110a-cに接続される送信先ノード104a-fを含む。入力ポート108に接続される発信元ノード102は、IPマルチキャストグループのためのIPマルチキャストデータパケットの発信元である。たとえば、発信元ノード102は、IPマルチキャストグループのすべてのメンバにビデオを伝送していてもよい。代替の実施例では、発信元ノード102は、IPマルチキャストグループのメンバであってもよい。発信元ノード102は、IPマルチキャストデータパケット106をスイッチ100に転送する。スイッチ100は、IPマルチキャストデータパケット106のコピーを、IPマルチキャストグループの各メンバ、すなわち送信先ノード104a-fに転送する。

#### 【0013】

IPマルチキャストデータパケット106のコピーは、各送信先ノード104

に転送される前に、スイッチ100によって修正され得る。受信されたIPマルチキャストデータパケット106のコピーは、スイッチ100に記憶される修正エントリに依存して修正される。受信されたIPマルチキャストデータパケットのコピーの修正および転送は、スイッチによってワイヤ速度で行なわれる、すなわち、修正は、ホストの介入なしに行なわれ、出力修正済IPマルチキャストデータパケット112は、送信先ノード104が接続される出力ポート110の次の使用可能ポートサイクルにおいて送信先ノード104に転送される。

#### 【0014】

IPマルチキャストデータパケット106が入力ポート108で受信されると、データパケット処理論理128が、IPマルチキャストデータパケット106のヘッダから情報を抽出し、抽出された情報を入力マルチキャスト転送論理114に転送する。入力マルチキャスト転送論理114は、IPマルチキャストグループ中の送信先ノード104がどの出力ポート110に接続されているかを判断し、IPマルチキャストデータパケットが転送されるべき出力ポート110を示す出力ポート-q論理134にマルチキャスト転送ベクトル120を転送する。入力マルチキャスト転送論理114はまた、IPマルチキャストデータパケット106が転送されるべきIPマルチキャストグループに対応するマルチキャスト転送インデックス（“MFI”）132を、データパケット処理論理128に転送する。MFI132は、IPマルチキャストグループの各メンバーのための修正エントリへのインデックスである。

#### 【0015】

データパケット処理論理128は、受信されたIPマルチキャストデータパケット106およびMFI132をメモリ116に転送する。MFI132を含む入力修正済IPマルチキャストデータパケット126は、メモリ116に記憶される。入力修正済IPマルチキャストデータパケット126のアドレスは、IPマルチキャストデータパケット106が転送されるべき各出力ポート110のためのポートキュー（図示せず）に記憶される。ポートキューは、デイビッド・エイ・ブラウン（David A. Brown）による「インタリーブされた非ブロッキングパケットバッファのための方法および装置」（“Method and Apparatus for an in

terleaved Non-Blocking Packet Buffer” ) と題する1999年8月31日に出版された同時係属中の米国出願連続番号第09/386, 589号に記載され、その全教示がここに引用により援用される。

#### 【0016】

入力修正済IPマルチキャストデータパケット126に記憶されたMFI132は、IPマルチキャストグループの各メンバと関連付けられる記憶された修正エントリにアクセスするために使用される。出力マルチキャスト転送論理118は、メンバのための記憶された修正エントリの内容を用いてメモリ116から受信された入力修正済IPマルチキャストデータパケット126のコピーを修正し、出力修正済IPマルチキャストデータパケット112をIPマルチキャストグループの各メンバに転送する。

#### 【0017】

IPマルチキャストグループの各メンバのための修正エントリを記憶することにより、IPマルチキャストデータパケット106のコピーを修正しIPマルチキャストグループのすべてのメンバに転送するのにホストの処理が必要とされない。したがって、IPマルチキャストデータパケットは、ワイヤ速度で、すなわち、ポートに接続されたIPマルチキャストグループのメンバにそのポートのための次の使用可能ポートサイクルにおいて転送される。各送信先ノード104と関連付けられる修正エントリは、出力マルチキャスト転送論理118に記憶され、入力マルチキャスト転送論理114から転送されたIPマルチキャストグループと関連付けられるMFI132を用いてアクセスされる。MFIは、メモリ116中の入力修正済マルチキャストデータパケット126に記憶される。

#### 【0018】

図1に示すように、IPマルチキャストグループのメンバは、送信先ノードDa-f104a-fを含む。送信先ノードDa-c104a-cは、出力ポート\_\_1 110aに接続され、送信先ノードDd-e104d-eは、出力\_\_ポート2 110bに接続され、送信先ノードDf104fは、出力ポート\_\_3 100cに接続される。図示の実施例におけるすべての送信先ノード104a-fは終端ノードであるが、代替の実施例においては、図示の送信先ノード104a

— f のいずれか1つが、別のスイッチの入力ポートであってもよく、IPマルチキャストデータパケットのコピーが、他のスイッチを介して、他のスイッチの出力ポートに接続されたIPマルチキャストグループのメンバに転送されてもよい。また、他のスイッチの出力ポートの出力ポートに接続される複数のVLANがあってもよく、各VLANが、IPマルチキャストデータパケットの修正されたコピーを要求する。

#### 【0019】

発信元ノードS a 1 0 2によって伝送されたデータパケット1 0 6は、入力ポート\_\_1 1 0 8でネットワークスイッチ1 0 0によって受信される。データパケット1 0 6は、宛先アドレスを含む。宛先アドレスがIPマルチキャストアドレスであるならば、スイッチ1 0 0は、IPマルチキャストデータパケット1 1 2の修正されたコピーを、スイッチ1 0 0に接続されたIPマルチキャストアドレスと関連付けられるIPマルチキャストグループのすべてのメンバに転送する。

#### 【0020】

図示のIPマルチキャストグループについては、スイッチ1 0 0は、6個の出力修正済IPマルチキャストデータパケット1 1 2 a — fを出力ポート1 — 3 1 1 0 a — cに転送する。出力修正済IPマルチキャストデータパケット1 1 2 は、IPマルチキャストデータパケット1 0 6が入力ポート1 0 2で受信された後の、出力ポート1 1 0のための次の使用可能ポートサイクルにおいて出力ポート1 1 0に転送される。

#### 【0021】

3個の出力修正済IPマルチキャストデータパケット1 1 2 a — cが、出力ポート\_\_1 1 1 0 Aに、1つが送信先ノードD a — c 1 0 4 a — cの各々に対して転送される。2個の出力修正済IPマルチキャストデータパケット1 1 2 d — eが、出力\_\_1 1 1 0 bに、1つが送信先ノードD d — e 1 0 4 d — eの各々に対して転送される。1個の出力修正済IPマルチキャストデータパケット1 1 2 fが、出力ポート\_\_2 1 1 0 cに、送信先ノードD f 1 0 4 fに対して転送される。

## 【0022】

受信されたIPマルチキャストデータパケット106を識別すると、データパケット処理論理128は、IPマルチキャストデータパケット130のヘッダの一部を入力マルチキャスト転送論理114に転送する。入力マルチキャスト転送論理114は、IPマルチキャストデータパケット106のコピーが転送されるべき出力ポート\_\_1-3 110a-cを判断し、マルチキャスト転送ベクトル120を生成する。マルチキャスト転送ベクトル120は、出力ポート\_\_q論理134に転送される。入力マルチキャスト転送論理114は、IPマルチキャストデータパケット130のヘッダの一部に依存して、記憶されたマルチキャスト転送インデックス(“MFI”)132を選択し、MFI132をデータパケット処理論理128に転送する。データパケット処理論理128は、受信されたIPマルチキャストデータパケット106とともにMFI132を入力修正済IPマルチキャストデータパケット126に記憶し、入力修正済IPマルチキャストデータパケット126をデータ\_\_in122によりメモリ116に転送する。入力修正済IPマルチキャストデータパケット126は、メモリ116に記憶される。

## 【0023】

出力ポート\_\_1-3 110a-cの各々について次のポートサイクルで、入力修正済IPマルチキャストデータパケット126は、出力ポート\_\_q論理134から転送されるアドレス136に依存して、メモリ116からデータ\_\_out124により出力マルチキャスト転送論理118に転送される。出力マルチキャスト転送論理118は、入力修正済データパケット126に含まれるMFI132に依存して入力修正済データパケット126のコピーを修正する。出力修正済データパケット112a-fは、それぞれの出力ポート\_\_1-3 110a-cに転送される。すべての出力修正済データパケット112a-fが転送されてしまった後、メモリ116に記憶された入力修正済データパケット126は上書きされてもよい。

## 【0024】

図2は、図1に示すIPマルチキャストグループのメンバへのスイッチ100



を介したIPマルチキャストデータパケット106のワイヤ速度転送を示すタイミング図である。図2は、図1と関連付けて記載される。図1に示すマルチキャストグループは、スイッチ100に接続された6個のメンバDa-Df104a-fを含む。スイッチ100は、4個のポート、入力\_\_ポート1 108および出力ポート1-3 110a-bを含む。しかしながら、この発明は、図示の4個のポートまたは6個のIPマルチキャストグループメンバに限られるものではない。図1の構成は、4個より多いポートや、6個より多いメンバを有するIPマルチキャストグループを含むように拡張されてもよいことが、当業者には周知である。

#### 【0025】

メモリ116は64バイト幅であるので、タイミング図において、入力ポート\_\_1 108で受信される各IPマルチキャストデータパケット106は、イーサネット(R)データパケットとして、たとえば64バイトの最小データパケットサイズであるものとする。しかしながら、この発明は、64バイトパケットサイズに限られるものでなく、データパケット中のバイトの数は、メモリ116の幅よりも大きくてもよい。

#### 【0026】

また、64バイトのIPマルチキャストデータパケット106は、データセグメントタイムスロット200の間に入力ポート108(図1)で受信されるものとする。図1に示すスイッチは、スイッチ100に接続された4個のポート、すなわち、1個の入力ポート108および3個の出力ポート110a-cを含むので、図示のデータセグメントタイムスロット200は、スイッチ100に接続された各ポートごとに1つの、4個のポートサイクルタイムスロット202を含む。スイッチ100によってサポートされるポートの最大数は、最小ポートサイクルタイムスロット202および最大データセグメントタイムスロット200に依存し、したがって図1に示す4個のポートに限られるものでない。4個より多いポートを備えるスイッチ100において、データセグメントタイムスロット200は、スイッチ100に接続された各ポートごとにポートサイクルタイムスロット202を含む。たとえば、イーサネット(R)データパケットについて、最小

データパケットサイズは64バイトである。したがって、最大データセグメントタイムスロット200は、入力ポートで64バイトを受信するのにかかる時間であり、最小ポートサイクルタイムスロット202は、メモリ116（図1）から64バイトにアクセスするための時間に依存する。データセグメントタイムスロットおよびポートサイクルタイムスロットは、デイビッド・エイ・ブラウンによる「インタリーブされた非ブロッキングパケットバッファ」と題する1999年8月31日に出願された同時係属中の米国出願連続番号第09/386,589号に記載され、その全教示がここに引用により援用される。

#### 【0027】

入力ポート\_\_1 108に到着するIPマルチキャストデータパケット106のコピーは、次の使用可能ポートサイクルタイムスロット202においてIPマルチキャストグループの各メンバに転送される、すなわち、それはワイヤ速度で転送される。IPマルチキャストグループの1つのみのメンバがポートに接続される場合、IPマルチキャストデータパケットのコピーは、次のポートサイクルタイムスロット202において転送される。IPマルチキャストグループの2つ以上のメンバがポートに接続される場合、入力ポート108に到着する次のIPマルチキャストデータパケットは、先のIPマルチキャストデータパケットがポートに接続されたIPマルチキャストグループのすべてのメンバに転送されてしまった後に、転送される。

#### 【0028】

タイミング図は、データセグメントタイムスロット200a-dにおいて発信元ノードa102から入力ポート\_\_1 108に到着する4個のIPマルチキャストデータパケットを示す。最初のIPマルチキャストデータパケットは、データセグメントタイムスロット200aにおいて到着する。第2のデータセグメントタイムスロット200bにおいて、IPマルチキャストデータパケットは、ポートサイクル202bにおいて出力ポート\_\_1 110aに、ポートサイクル202cにおいて出力ポート\_\_2 110bに、ポートサイクル202cにおいて出力ポート\_\_3 110cに転送される。最初のIPマルチキャストデータパケットが転送されている際、第2のIPマルチキャストデータパケットがデータセ

グメントタイムスロット200bの間に受信される。

【0029】

時間230において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) がメモリ116 (図1) に記憶され、入力マルチキャスト転送論理114 (図1) は、IPマルチキャストデータパケット\_\_1 106 (図1) のためのマルチキャスト転送ベクトル120 (図1) を出力マルチキャスト転送論理118 (図1) に転送する。

【0030】

時間206、ポートサイクルタイムスロット202bの開始において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) のコピーが、メモリ116 (図1) からデーターout124 (図1) により出力マルチキャスト転送論理118 (図1) へ、出力ポート\_\_q 論理134 (図1) から転送されたアドレス136に依存して、転送される。出力マルチキャスト転送論理118 (図1) は、MFI132 (図1) に依存して入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) の受信されたコピーを修正する。出力修正済IPマルチキャストデータパケット\_\_1 112Aは、出力ポート\_\_1 110Aを介して送信先ノードDa104aに転送される。

【0031】

時間208、ポートサイクルタイムスロット202cの開始において、および、時間210、ポートサイクルタイムスロット202dの開始において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) のコピーがメモリ116 (図1) からデーターout124 (図1) により、出力ポート\_\_q 論理134 (図1) から転送されたアドレス136 (図1) に依存して、転送される。入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) の受信されたコピーは、関連付けられたMFI132 (図1) に依存して出力マルチキャスト転送論理118によって修正され、出力ポート2 110bを介して送信先ノードDd104dおよび出力ポート3 110cを介して送信先ノードDf104fに転送される。

【0032】

時間212において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) のコピーが修正され送信先ノードDb 104b (図1) に転送され、時間218において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) のコピーが修正され、出力ポート\_\_2 110b (図1) に接続された送信先ノードDc 104c に転送される。時間214において、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) のコピーが修正され、出力ポート\_\_1 110b (図1) に接続された送信先ノードDd 104d に転送される。

### 【0033】

時間232において、入力修正済IPマルチキャストデータパケット\_\_2 126 (図1) がメモリ116 (図1) に記憶され、入力マルチキャスト転送論理114 (図1) は、IPマルチキャストデータパケット\_\_2 106 (図1) のためのマルチキャスト転送ベクトル120 (図1) を出力マルチキャスト転送論理118 (図1) に転送する。時間234において、入力修正済IPマルチキャストデータパケット\_\_3 126 (図1) がメモリ116 (図1) に記憶される。時間236において、入力修正済IPマルチキャストデータパケット\_\_4 126 (図1) がメモリ116 (図1) に記憶される。

### 【0034】

IPマルチキャストグループの1つのみのメンバ、すなわち、送信先ノード140fは出力ポート\_\_3 110cに接続されているので、1つのみの出力修正済IPマルチキャストデータパケット\_\_1 112 (図1) が出力ポート\_\_3 110cに転送される。したがって、時間216において、ポートサイクルタイムスロット202dにおいて、出力修正済IPマルチキャストデータパケット\_\_2 112fが出力ポート\_\_3 110Cに転送される。時間222で、ポートタイムサイクル202dにおいて、出力修正済IPマルチキャストデータパケット\_\_3 112fが出力ポート\_\_3 110cに転送される。時間228で、ポートタイムサイクルタイムスロット3 202dにおいて、出力修正済IPマルチキャストデータパケット\_\_4 112fが出力ポート\_\_3 110c (図1) に転送される。

## 【0035】

したがって、入力ポート\_\_0 108 (図1) で受信されたIPマルチキャストデータパケット\_\_1 106 (図1) は、スイッチ100 (図1) を介して、次の使用可能ポートサイクルタイムスロット202においてホスト処理なしにワイヤ速度で、出力ポート\_\_1-3 110a-c (図1) に接続された送信先ノード104a-f (図1) に転送される。

## 【0036】

出力修正済IPマルチキャストデータパケット\_\_2 112 (図1) は、出力ポート\_\_2 110bにより送信先ノードDe104eおよび送信先ノードDf104fに転送される。時間220で、出力修正済IPマルチキャストデータパケット\_\_2 112dが、出力ポート\_\_2 110bに接続された送信先ノードDd104dに転送される。時間226で、出力修正済IPマルチキャストデータパケット\_\_2 112e (図1) が、出力ポート\_\_2 110bに接続された送信先ノードDe104eに転送される。

## 【0037】

時間224で、最後の出力修正済IPマルチキャストデータパケット\_\_1 112 (図1) が転送されてしまった後、入力修正済IPマルチキャストデータパケット\_\_1 126 (図1) を記憶するメモリ中の場所を用いて、入力ポート108 (図1) で受信された別のデータパケットを記憶し得る。したがって、入力ポート108 (図1) で受信された入力修正済IPマルチキャストデータパケット126 (図1) は、出力修正済IPマルチキャストデータパケット112 (図1) がIPマルチキャストグループのすべてのメンバに転送されてしまうまでメモリ116 (図1) に記憶される。

## 【0038】

すべての出力修正済IPマルチキャストデータパケット112a-f (図1) が転送されてしまったかどうかを判断するために、入力修正済IPマルチキャストデータパケット126 (図1) が、各入力修正済IPマルチキャストデータパケット126 (図1) に関連付けられて、入れられたポートキュー (図示せず) の数が、ポートキューカウンタレジスタ (図示せず) に記憶される。出力ポート

110 (図1) と関連付けられるすべての出力修正済IPマルチキャストデータパケット112 (図1) が出力ポート110 (図1) に転送されるたびに、カウンタレジスタがデクリメントされる。たとえば、1個の出力修正済IPマルチキャストデータパケットが出力ポート\_\_3 110cに転送されてしまった後に、ポートキューのカウンタがデクリメントされる。ポートキューの数は、マルチキャスト転送ベクトル120 (図1) から判断される。

#### 【0039】

入力ポート108 (図1) で受信される先行技術のデータパケットが、出力修正済IPマルチキャストデータパケット112a-f (図1) が出力ポート110a-cに転送される前に転送論理114、118 (図1) およびデータパケット処理論理126 (図1) がIPマルチキャストパケット106 (図1) に対して行ない得る修正を説明するために、図3A-3Dと関連付けて記載される。

#### 【0040】

図3Aは、スイッチ100 (図1) に接続された入力ポート108 (図1) で受信され得る先行技術のデータパケットを示す。図3Bは、図3Aに示すデータパケットに含まれ得る先行技術のイーサネット (R) ヘッダを示す。図3Cは、図3Aに示すデータパケットに含まれ得る先行技術のインターネットプロトコルヘッダを示す。図3Dは、図3Bに示すイーサネット (R) ヘッダ中の先行技術のタグ制御情報フィールド318Bを示す。

#### 【0041】

図3Aは、先行技術のデータパケット300を示す。データパケット300は、データフィールド310と、ネットワーク層のためのヘッダ302、304、306、308とを含む。OSIモデルにおける4個の層のためのヘッダ、すなわち、物理層 (L1) ヘッダ302と、データリンク層 (L2) ヘッダ304と、ネットワーク層 (L3) ヘッダ306と、トランスポート層 (L4) ヘッダ308とが示される。たとえば、データリンク層 (L2) ヘッダ304はイーサネット (R) であってもよく、ネットワークワーキング層 (L3) ヘッダ306はインターネットプロトコル (“IP”) であってもよい。データパケット300は、チェックサム312も含む。

**【0042】**

図3Bは、イーサネット(R) データリンク(L2) ヘッダ304のフォーマットを示す。イーサネット(R) データリンク(L2) ヘッダ304は、送信先ノード104(図1)のための装置アドレス、すなわち、宛先アドレス314と、発信元ノード102(図1)のための装置アドレス、すなわち、出所アドレス316と、オプションの仮想ローカルエリアネットワーク識別子(“VLAN ID”) フィールド318と、長さ/タイプフィールド320とを含む。VLAN ID318は、タグプロトコル識別子(“TPID”) フィールド318Aと、タグ制御情報(“TCI”) フィールド318Bとを含む。VLAN IDフィールド318は、IEEE802.1Qタグ付けおよびIEEE802.1D1988(802.1p) 優先権レベルに基づいてVLANスイッチングのためのサポートを提供する。

**【0043】**

図3Cは、IPネットワーク層(L3) ヘッダ306のフォーマットを示す。IPネットワーク層(L3) ヘッダ306は、発信元ノード102(図1)のためのネットワークアドレス、すなわち出所IPアドレス344と、送信先ノード104(図1)のためのネットワークアドレス、すなわちIP宛先アドレス346とを含む。IPネットワーク層ヘッダ306中の他のフィールドは、バージョン322と、HLEN324と、サービスタイプ(“TOS”) 326と、パケット長(total length) 328と、識別子330と、フラグ332と、フラグメントオフセット334と、生存時間(time to live) (“TTL”) 336と、プロトコル340と、ヘッダチェックサム342と、オプション348と、パッド350とを含む。

**【0044】**

スイッチ100(図1)に接続された入力ポート108(図1)で受信されたIPマルチキャストデータパケット106(図1)は、イーサネット(R) データリンク(L2) ヘッダ300(図3A)に記憶された宛先アドレス314(図3B)に依存して1つ以上の出力ポート110a-c(図1)にブリッジされるか、または、IPネットワーク層(L3) ヘッダ306に記憶されたIP宛先ア

ドレス346に依存して1つ以上の出力ポート110（図1）にルーチングされる。

#### 【0045】

図3Dは、図3Bに示す先行技術のTCIフィールド318bを示す。TCIフィールド318bは、3ビットのユーザ優先権フィールド352と、1ビットのCFIフィールド354と、12ビットの仮想ローカルエリアネットワーク（“VLAN”）識別子356とを含む。

#### 【0046】

図1に戻って、データパケット処理論理128は、受信されたデータパケットに含まれるネットワーク層（L3）ヘッダ306（図3A）を調べることにより、入力ポート\_\_1 108で受信されたデータパケットがIPマルチキャストデータパケット106であるかどうかを判断する。ネットワーク層（L3）ヘッダ306（図3A）は、プロトコルフィールド340（図3C）がIPであり、バージョンフィールド322（図3C）がIPv4であり、かつヘッダチェックサム342（図3C）が有効であるならば、インターネットプロトコルバージョン4（“IPv4”）ヘッダである。IPデータパケットは、IP宛先アドレス346（図3C）がクラスDであるならば、すなわち、IP宛先アドレス346（図3C）の上位4ビットが1.110にセットされているならば、IPマルチキャストデータパケット106である。

#### 【0047】

データパケット処理論理128がデータパケットがIPマルチキャストデータパケット106であると判断したならば、IPネットワーキング層（L3）ヘッダ306（図3C）中のIP宛先アドレス346（図3C）およびIP出所アドレス344（図3C）、イーサネット（R）データリンク（L2）ヘッダ304（図3B）中のVLAN ID318（図3B）、ならびに入力ポート108（図1）に対応する入力ポート番号を含むIPマルチキャストデータパケット130のヘッダの一部が、入力マルチキャスト転送論理114に転送される。

#### 【0048】

図4は、図1に示す入力マルチキャスト転送論理114を示す。入力マルチキ



キャスト転送論理114は、内容アドレス記憶装置（“CAM”）400と、ルートマップ402と、ルートテーブル404と、ベクトル組合せ論理406とを含む。ルートテーブル404は、入力ポート\_\_1 108（図1）で受信されるIPマルチキャストデータパケット106（図1）に含まれるIPマルチキャストグループに対応するマルチキャスト転送エントリ422を含む。CAM400およびルートマップ402は、データパケット処理論理128（図1）から転送されるIPマルチキャストデータパケット130（図1）のヘッダの一部および入力ポート\_\_1 108（図1）のポート番号に依存して、マルチキャスト転送エントリ422を選択する。

#### 【0049】

図4と関連付けて記載される入力マルチキャスト転送論理114（図1）の実現において、CAM400は、IP宛先アドレス346（図3C）、IP出所アドレス344（図3C）、入力ポート番号およびVLAN ID318（図3B）を記憶するには十分な幅がない。したがって、ルートマップ402は、CAM400の幅を増大させるように実現される。しかしながら、CAM400がルートマップ402を含むよう拡張されてもよいことは、当業者には周知である。

#### 【0050】

IPマルチキャストパケット106に対応するルートテーブル404におけるマルチキャスト転送エントリ422の検索は、2段階で行なわれる。まず、IPネットワーク層（L3）ヘッダ306（図3C）中のIP宛先アドレス346（図3C）フィールドおよびIP出所アドレス344（図3C）フィールドに依存してエントリのためにCAM400が検索される。一致するエントリがCAM400に見出されるならば、一致アドレス410がルートマップ402に転送される。ルートマップ402は、一致アドレス410に対応するマルチキャストパスコード（“MPC”）422をCAM400に転送する。

#### 【0051】

第2に、MPC422、VLAN ID318（図3B）および入力ポート番号に依存して、エントリのためにCAM400が検索される。一致するエントリがCAM400に見出されるならば、一致アドレス410がルートマップ402

に転送される。ルートマップ402は、一致アドレス410に対応する関連データインデックス(“ADI”)をルートテーブル404に転送する。IPマルチキャスト転送エントリ422は、ADI412によって識別されるルートテーブル404中の場所に記憶される。IPマルチキャスト転送エントリ422は、IPマルチキャストデータパケット106(図1)がスイッチ100(図1)によって転送されるべきIPマルチキャストグループを識別する。

#### 【0052】

IPマルチキャストデータパケットの二重転送を防ぐために、CAM400は、IP出所アドレス344(図3C)、IP宛先アドレス346(図3C)、VLAN ID318(図3B)および入力ポート番号108(図1)で検索される。たとえば、発信元ノードが2つ以上のポートに接続されている場合、すなわち、送信先ノードxがポートyおよびポートxに接続され、かつIPマルチキャストデータパケットが送信先ノードxからポートyで受信されるならば、データパケットは、IPマルチキャストグループのすべてのメンバに転送される。しかしながら、IPマルチキャストデータパケットがポートzの宛先アドレスxから受信されるならば、それは廃棄される。なぜなら宛先xは2つ以上のポートに接続され、そのためポートzで受信されるIPマルチキャストデータパケットは、ポートyで受信されたIPマルチキャストデータパケットの複製であるからである。

#### 【0053】

ルートテーブル404中の各マルチキャスト転送エントリ422は、ブリッジベクトル416と、ルートベクトル414と、マルチキャスト転送インデックス132(“MFI”)とを含む。ブリッジベクトル416およびルートベクトル414は、ベクトル組合せ論理406に転送される。MFI132は、データパケット処理論理128(図1)に転送される。

#### 【0054】

ベクトル組合せ論理406は、ルートベクトル414およびブリッジベクトル416に対して論理OR関数を実行することによりマルチキャスト転送ベクトル120を生成する。マルチキャスト転送ベクトル120は、出力ポート\_\_q論理

134 (図1) に転送される。マルチキャスト転送ベクトル120は、IPマルチキャストデータパケット (図1) が転送されるべき1つ以上の出力ポートを選択する。

#### 【0055】

入力ポート\_\_1 108で受信されるIPマルチキャストデータパケット106 (図1) のイーサネット (R) データリンク (L2) ヘッダ304 (図3B) における宛先アドレス314 (図3B) の最初の3バイトは、“0x01-00-53” にセットされる。IPマルチキャストデータパケット106がデータイン122 (図1) によりメモリ116 (図1) に転送される前に、データパケット処理論理128 (図1) は、入力マルチキャスト論理114 (図1) から転送されたMFI132を宛先アドレス314 (図3B) の第2および第3のバイトに以下のとおり書込む：“0x01-0M-F1”。入力修正済IPマルチキャストデータパケット126 (図1) はメモリ116 (図1) に記憶される。出力マルチキャスト転送論理118 (図1) は、入力修正済IPマルチキャストデータパケット126 (図1) に含まれるMFI132を用いて、出力修正済IPマルチキャストデータパケット112 (図1) を出力ポート110 (図1) に転送する前に、入力修正済マルチキャストデータパケット126 (図1) を修正する。

#### 【0056】

IPマルチキャスト転送ベクトル120は、IPマルチキャストデータパケット106 (図1) がどの出力ポート110a-c (図1) に転送されるべきかを判断する。メモリ116 (図1) 中の入力修正済マルチキャストデータパケット126 (図1) の場所は、マルチキャスト転送ベクトル420において図5に示す入力マルチキャスト転送論理114で実現される対応の出力ポートキュービット (図示せず) の状態に依存して、出力ポート\_\_q 論理134 (図1) 中の出力ポートキュー (図示せず) に記憶される。

#### 【0057】

図5は、入力ポート108 (図1) で受信されるIPマルチキャストデータパケット106 (図1) を処理するための、図4に示す入力マルチキャスト転送論

理114において実現される方法のフローチャートである。この方法は、図1－図4と関連付けて記載される。

【0058】

ステップ500で、入力マルチキャスト転送論理114（図1）は、データパケット処理論理128（図1）からIPマルチキャストデータパケット130の一部を受信する。処理はステップ502に続く。

【0059】

ステップ502で、入力マルチキャスト転送論理114（図4）は、ベクトル組合せ論理406（図4）においてルートベクトル414（図4）およびブリッジベクトル416（図4）を初期化する。

【0060】

ステップ504で、データパケット処理論理128（図1）からCAM400（図4）に転送されるIP出所アドレス344（図3C）とIP宛先アドレス346（図3C）とを一致させるエントリのためにCAM400（図4）において検索が行なわれる。一致アドレス410（図4）がCAM400（図4）に見出されるならば、処理はステップ506に続く。そうでなければ、処理はステップ518に続く。

【0061】

ステップ506で、一致アドレス410（図4）は、ルートマップ402（図4）に転送される。ルートマップ402（図4）は、一致アドレス410（図4）に対応するマルチキャストパスコード（“MPC”）422（図4）をCAM400（図4）に転送する。処理はステップ508に続く。

【0062】

ステップ508で、MPC422と、イーサネット（R）データリンク（L2）ヘッダ304（図3B）からのVLAN ID318（図3B）と、入力ポート番号とに依存して、エントリのためにCAM400（図4）において第2の検索が行なわれる。一致アドレス410（図4）がCAM400（図4）に見出されるならば、処理はステップ510に続く。そうでなければ、処理はステップ518に続く。

**【0063】**

ステップ510で、一致アドレス410（図4）がルートマップ402（図4）に転送される。ルートマップ402（図4）は、一致アドレス410（図4）に対応するADIをルートテーブル404に転送する。IPマルチキャストデータパケット106（図1）のためのIPマルチキャスト転送エントリ422（図4）は、ADI412によって識別されるルートテーブル404中の場所に記憶される。処理はステップ512に続く。

**【0064】**

ステップ512で、マルチキャスト転送エントリ422（図4）に記憶されたMFI132（図4）が、データパケット処理論理128（図1）に転送され、マルチキャスト転送エントリ422（図4）に記憶されたブリッジベクトル416が、ベクトル組合せ論理406（図4）に転送される。処理はステップ514に続く。

**【0065】**

ステップ514で、IPマルチキャストデータパケット130（図1）の一部上で入力マルチキャスト転送論理114（図1）に転送されるIPマルチキャストデータパケット106（図1）中のIPネットワーク層（L3）ヘッダ306（図3C）中のTTLフィールド336（図3C）が、1より大きければ、処理はステップ516に続く。そうでなければ、処理はステップ518に続く。

**【0066】**

ステップ516で、マルチキャスト転送エントリ422（図4）に記憶されたルートベクトル414（図4）は、ベクトル組合せ論理406（図4）に転送される。処理はステップ518に続く。

**【0067】**

ステップ518で、IPマルチキャストデータパケット106（図1）が受信された入力ポート108（図1）に対応する入力ポート番号がブリッジベクトル416（図4）から除去される。処理はステップ519に続く。

**【0068】**

ステップ519で、ブリッジベクトル416（図4）とVLANメンバシップ

ベクトル（図示せず）とが組合される。処理はステップ520に続く。

【0069】

ステップ520で、論理OR関数を実行することによりブリッジベクトル416（図4）とルートベクトル414（図4）とを組合せて、IPマルチキャストグループのためのマルチキャスト転送ベクトル120（図4）を与える。処理はステップ522に続く。

【0070】

ステップ522で、マルチキャスト転送ベクトル422は、出力ポート\_\_q論理134（図1）に転送される。

【0071】

IPマルチキャストデータパケット106（図1）に対してCAM400（図4）において一致するエントリが見出されなければ、すなわちIPマルチキャストグループがわからなければ、ホストポートがCAM400（図4）、ルートマップ402（図6）およびルートテーブル404（図4）においてIPマルチキャストグループのためのエントリを加え得るように、マルチキャストデータパケット106（図1）はホストポート（図示せず）に転送される。IPマルチキャストパケット106がホストポートに転送される（図1）ように、マルチキャスト転送ベクトル422（図4）は選択される。わからないIPマルチキャストグループのためのIPマルチキャストデータパケット106（図1）が、IPマルチキャストデータパケット106（図1）が受信されたVLANにブリッジされるようにも、マルチキャスト転送ベクトル422（図4）は選択される。

【0072】

図6は、図1に示す出力マルチキャスト転送論理118を示す。出力マルチキャスト転送論理118（図1）は、MFI検出論理602と、エディタ論理600と、MFIテーブル604とを含む。MFIテーブル604は、MFIエントリ612を含む。

【0073】

データーout124（図1）により転送されメモリ116（図1）に記憶された入力修正済IPマルチキャストデータパケット126（図1）のコピーが出

力IPマルチキャスト処理論理118によって受信されると、MFI検出論理602は、入力修正済IPマルチキャストデータパケット126（図1）に含まれるMFI132を検出し、MFI132をMFIテーブル604に転送する。MFI132は、IPマルチキャストグループに対応するMFIテーブル604中のMFIエン트리612へのインデックスである。MFIエン트리612は、IPマルチキャストデータパケット106（図1）の最初のコピーがどの出力ポート112（図1）に転送されるべきかを判断する。その後のIPマルチキャストデータパケット106が出力ポート110に転送されるべきならば、次のMFI132'は、MFIテーブル604中のMFIエン트리612に含まれる。次のMFI132'は、次のMFIエン트리612を選択するために、MFI検出論理602に転送される。たとえば、次のMFI132'は、出力修正済IPマルチキャストデータパケット112b（図1）のための次のMFIエン트리112'を選択するために、出力修正済IPマルチキャストデータパケット112a（図1）のためのMFIエン트리612とともに記憶される。

#### 【0074】

入力修正済IPマルチキャストデータパケット126（図1）は、MFIエン트리612に含まれるマルチキャストパラメータ610に依存してエディタ論理600によって修正される。MFIエン트리612に記憶されるマルチキャストパラメータ610は、エディタ論理600に転送される。したがって、各入力修正済IPマルチキャストデータパケット126（図1）は、それが転送される送信先ノード104a-f（図1）に依存して編集される。MFIテーブル604中にMFIエン트리612を設けることにより、入力修正済IPマルチキャストデータパケット126（図1）は、出力修正済IPマルチキャストデータパケット112（図1）を与えるように編集され、ワイヤ速度で、出力ポート110a-c（図1）に接続されるIPマルチキャストグループの各メンバに転送され得る。

#### 【0075】

図7は、図1に示す4個のポートを備えるスイッチ100に対する図6に示すMFIテーブル604を示す。MFIテーブル604（図6）は、IVIDテー

ブル700と、VLAN IDテーブル702とを含む。MFIエントリ612'は、IVIDテーブル700にIVIDエントリ718と、VLAN IDテーブル702にVLAN IDエントリ710とを含む。IVIDテーブル700中のIVIDエントリ718は、スイッチ100 (図1) に接続される各ポートごとにポートフィールド714 a-dを含む。図1に示すスイッチ100は、4個のポート：入力ポート\_\_1 108および出力ポート\_\_1-3 110 a-cを含む。

#### 【0076】

各ポートフィールド714は、ブリッジ/ルートフィールド708と、次のフィールド706と、IVIDフィールド704とを含む。IVIDエントリ718は、次のMFIフィールド716も含む。IVIDエントリ718中のいずれかの次のフィールド706 a-dが、別の出力修正済IPマルチキャストデータパケット112 (図1) がスイッチ100 (図1) に接続されるポートに転送されるべきことを示すならば、次のMFIフィールドは、次のMFI132'を記憶する。図示のIVIDエントリ718は、図1に示すように、4個のポートを備えるスイッチ100 (図1) のためのものである。ポートフィールド714の数は、図示の4個に限られるものでなく、ポートフィールド714は、スイッチ100 (図1) に接続される各ポートごとに設けられる。たとえば、64個のポートを備えるスイッチに対して、IVIDエントリ718は、64個のポートフィールド714を含む。

#### 【0077】

ポートに対する有効なVLAN IDエントリ716があるならば、IVIDフィールド704は、IVIDフィールド704に対応するVLAN IDテーブル702中のVLAN IDエントリ716へのインデックスを記憶する。VLAN IDテーブル702中のVLAN IDエントリ716は、VLAN IDフィールド710と、タグフィールド712とを含む。

#### 【0078】

代替の実施例では、IVIDテーブル700中のIVIDエントリ718は、VLAN IDエントリ716に記憶されたVLAN IDフィールド710お



よびタグフィールド712を含んでもよく、この場合にはIVIDフィールド704が必要とされないであろう。しかしながら、IVIDエントリ718およびVLAN IDエントリ710を、別個に、すなわち、IVIDテーブル700とVLAN IDテーブル702とに記憶することは、MFIテーブル604（図6）を実現するのに必要とされる総メモリを低減する、なぜなら、あらゆる入力修正済IPマルチキャストデータパケット126が、対応のVLAN IDエントリ716を有するとは限らないからである。

#### 【0079】

ブリッジ／ルートフィールド708は、単一ビットとして実現される。ブリッジ／ルートフィールド708の状態は、入力修正済IPマルチキャストデータパケット126（図1）がポートにブリッジまたはルーティングされるべきかどうかを判断する。次のフィールド706は、単一ビットとして実現される。次のフィールド706の状態は、別の出力修正済IPマルチキャストデータパケット112（図1）がポートに転送されるべきかどうかを判断する。IVIDフィールド704は、VLAN IDテーブル702へのインデックスを提供する。IVIDフィールド704中のビットの数は、VLAN IDテーブル702のサイズに依存する。たとえば、256個のVLAN IDエントリ716を備えるVLAN IDテーブル702に対しては、VLAN IDインデックス714は8ビットを含む。

#### 【0080】

入力修正済IPマルチキャストデータパケット126（図1）がポートにルーティングされるべきならば、VLAN ID710は、有効なVLAN ID710を含む。タグフィールド712が、出力修正済IPマルチキャストデータパケット112（図1）がタグ付けされていないことを示すならば、エディタ論理600（図6）は、イーサネット（R）データリンク層（L2）ヘッダ304（図3B）中のVLAN ID318（図3B）からタグ318b（図3B）を削除する。

#### 【0081】

VLAN ID710が必要とされることを判断し、VLAN ID710を

VLAN IDテーブル702から獲得すると、マルチキャストパラメータ610 (図6) がエディタ論理600 (図6) に転送される。エディタ論理600 (図6) は、マルチキャストパラメータ610を用いて、入力修正済IPマルチキャストデータパケット126 (図6) を修正し、修正した後、エディタ論理600 (図6) は、出力修正済IPマルチキャストデータパケット112を出力ポート110 (図1) に転送する。

#### 【0082】

ポートフィールド714におけるブリッジ/ルートフィールド708がブリッジにセットされたならば、エディタ論理600 (図6) は、VLAN ID318 (図3C) を付加するかまたは除去し、イーサネット (R) データリンク層 (L2) ヘッダ304 (図3B) 中の宛先アドレス314 (図3B) をその元の値に戻す。

#### 【0083】

ポートフィールド714中のブリッジ/ルートフィールド708がルーチングにセットされるならば、エディタ論理600 (図6) は、イーサネット (R) データリンク層 (L2) ヘッダ304 (図3B) における出所アドレス314 (図3B) にルータの出所アドレスを上書きし、TTL336 (図3C) をデクリメントし、IPチェックサムを生成してそれをヘッダチェックサム342 (図3C) に記憶する。エディタ論理600 (図6) はまた、IP宛先アドレス346 (図3C) を元の値に戻し、TCI318b (図3B) タグを付加する。

#### 【0084】

出力マルチキャスト転送論理118 (図1) は、IP宛先アドレス346 (図3C) に、出力ポート110a-c (図1) への出力修正済IPマルチキャストデータパケット112a-f (図1) を転送する。

#### 【0085】

この発明は、その好ましい実施例を参照して特定の示され記載されたが、前掲の特許請求の範囲によって含まれるこの発明の範囲から逸脱することなしに態様および詳細においてさまざまな変更がこれになされ得ることが、当業者には理解されるであろう。

**【図面の簡単な説明】**

【図 1】 この発明の原理に従う、入力ポートで受信された I P マルチキャスト packets をスイッチに接続された I P マルチキャストグループのメンバに転送するための I P マルチキャスト転送論理を含むネットワークスイッチの図である。

【図 2】 図 1 に示すスイッチによる、図 1 に示す I P マルチキャストグループのメンバへの I P マルチキャストデータパケットの転送を示すタイミング図である。

【図 3 A】 スwitch に接続された入力ポートで受信され得る先行技術のデータパケットの図である。

【図 3 B】 図 3 A に示すデータパケットに含まれ得る先行技術のイーサネット (R) データリンク層 (L 2) ヘッダの図である。

【図 3 C】 図 3 A に示すデータパケットに含まれ得る先行技術のインターネットプロトコル (ネットワーク層 (L 3)) ヘッダの図である。

【図 3 D】 図 3 B に示す先行技術のタグ制御情報の図である。

【図 4】 図 1 に示す入力 I P マルチキャスト転送論理の図である。

【図 5】 入力ポートで受信される I P マルチキャストデータパケットを処理するための方法の流れ図である。

【図 6】 図 1 に示す出力マルチキャスト転送論理の図である。

【図 7】 図 6 に示すマルチキャスト転送インデックステーブルの図である。

。

【図 8】 I P マルチキャストデータパケットを修正するための、図 6 に示す出力マルチキャスト転送論理において行なわれるステップの流れ図である。

【図1】

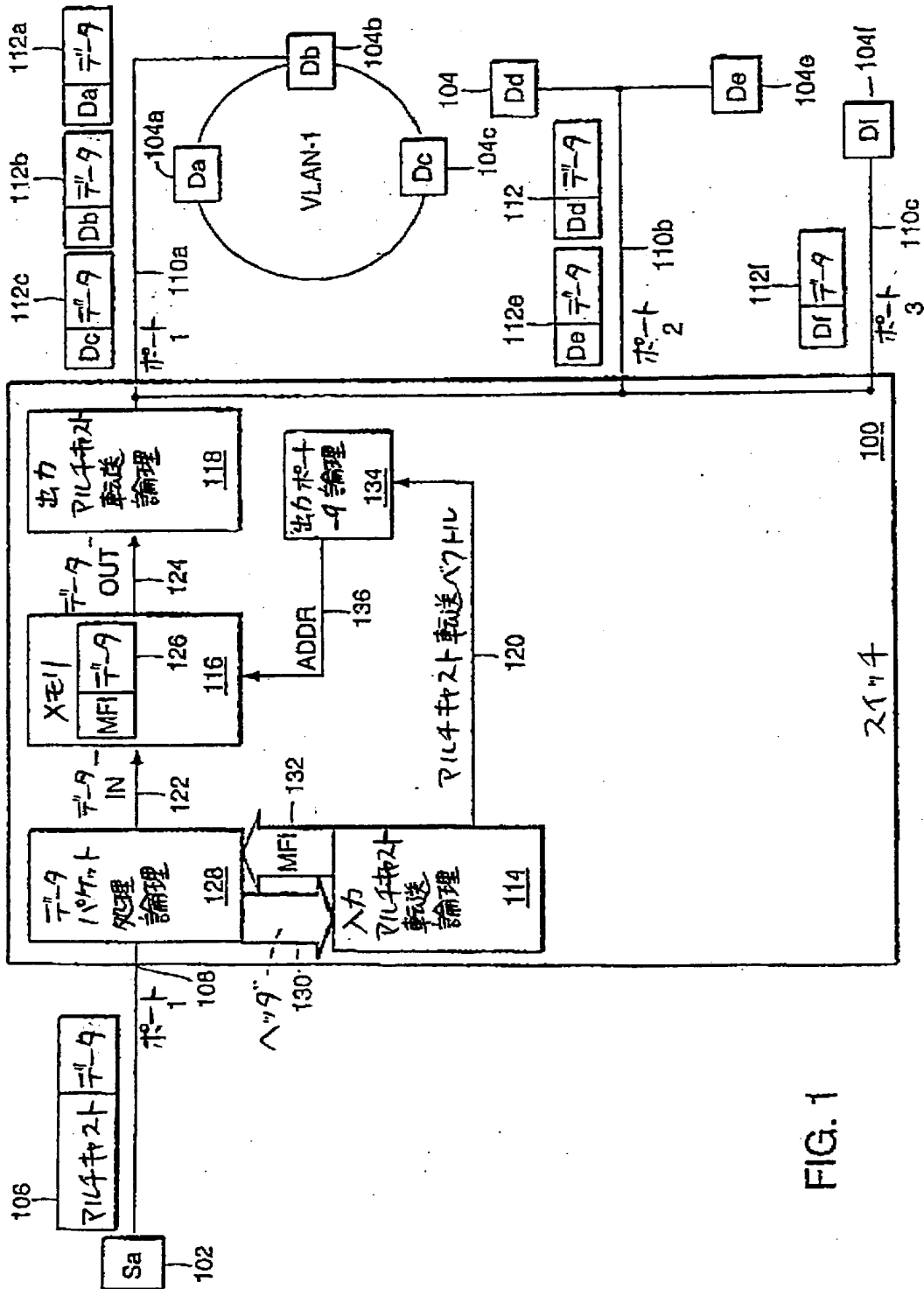


FIG. 1

【図2】

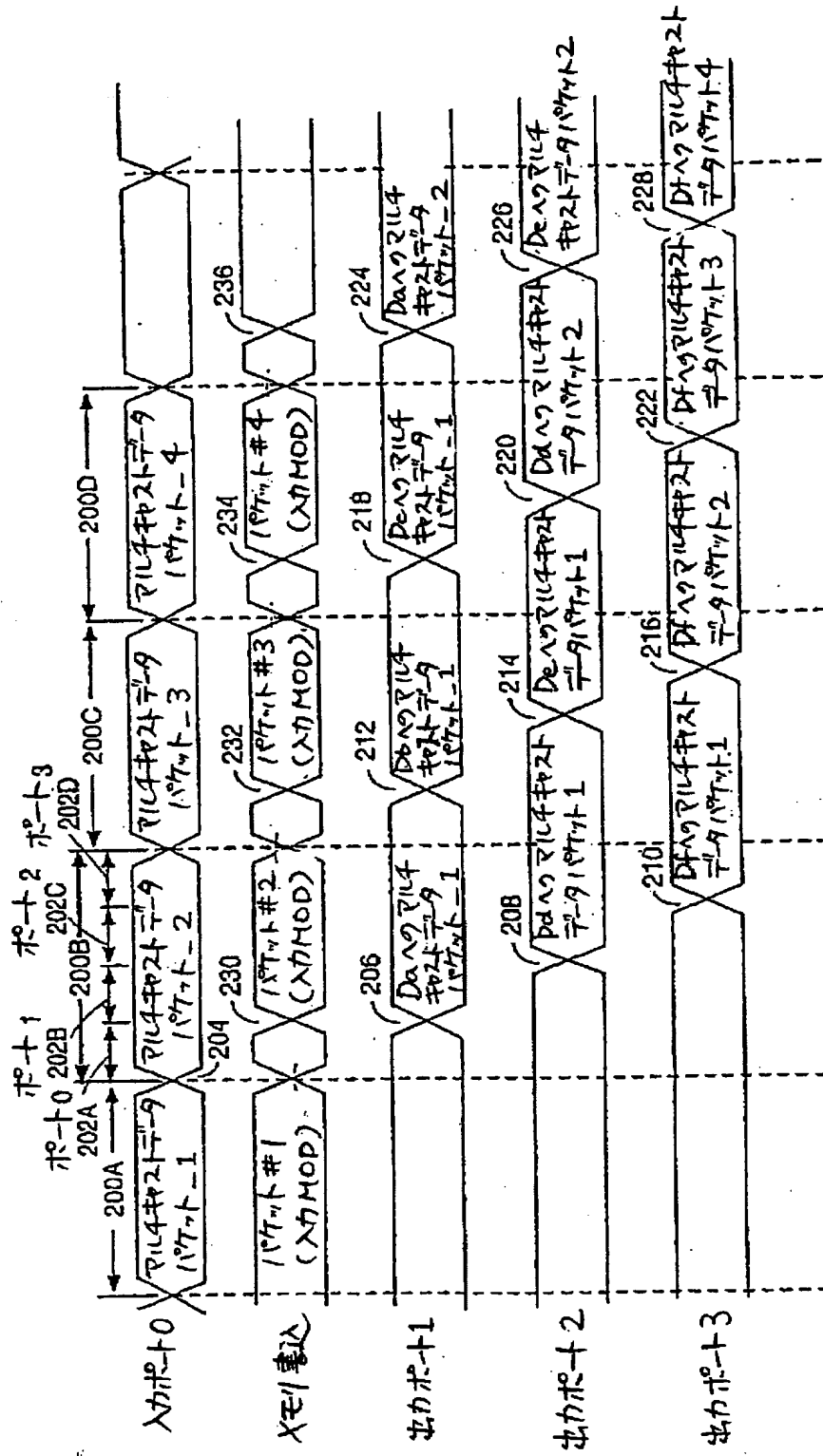


FIG. 2

【図 3A】

300	物理層 (L1) ヘッダ 302	データリンク層 (L2) ヘッダ 304	ネットワーク層 (L3) ヘッダ 306	トランスポート層 (L4) ヘッダ 308	データ 310	アプリケーション 312
-----	---------------------------	-------------------------------	-------------------------------	--------------------------------	------------	-----------------

PRIOR ART  
FIG. 3A

【図3C】

306

<u>322</u> VERS	<u>324</u> HLEN	TOS <u>326</u>	<u>328</u> 1バイト長
<u>330</u> 識別子		<u>332</u> フラグ	フラグメント オフセット <u>334</u>
<u>336</u> TTL	<u>340</u> プロトコル	<u>342</u> ヘッダチェックサム	
SRC IP アドレス			<u>344</u>
DEST IP アドレス			<u>346</u>
<u>348</u> オプション			<u>350</u> パッド

PRIOR ART  
FIG. 3C

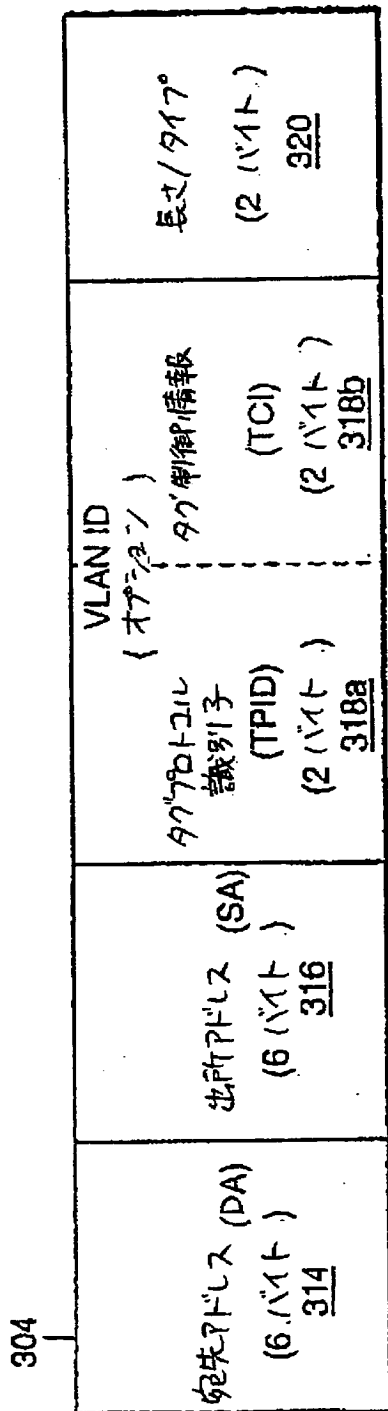
【図3D】

318B

ユーザ 優先権 (3ビット) <u>352</u>	<u>354</u> CFI (1 ビット)	<u>356</u> VLAN 識別子 (12ビット)
------------------------------------	---------------------------------	-----------------------------------

PRIOR ART  
FIG. 3D

【図 3 B】



PRIOR ART  
FIG. 3B



【図5】

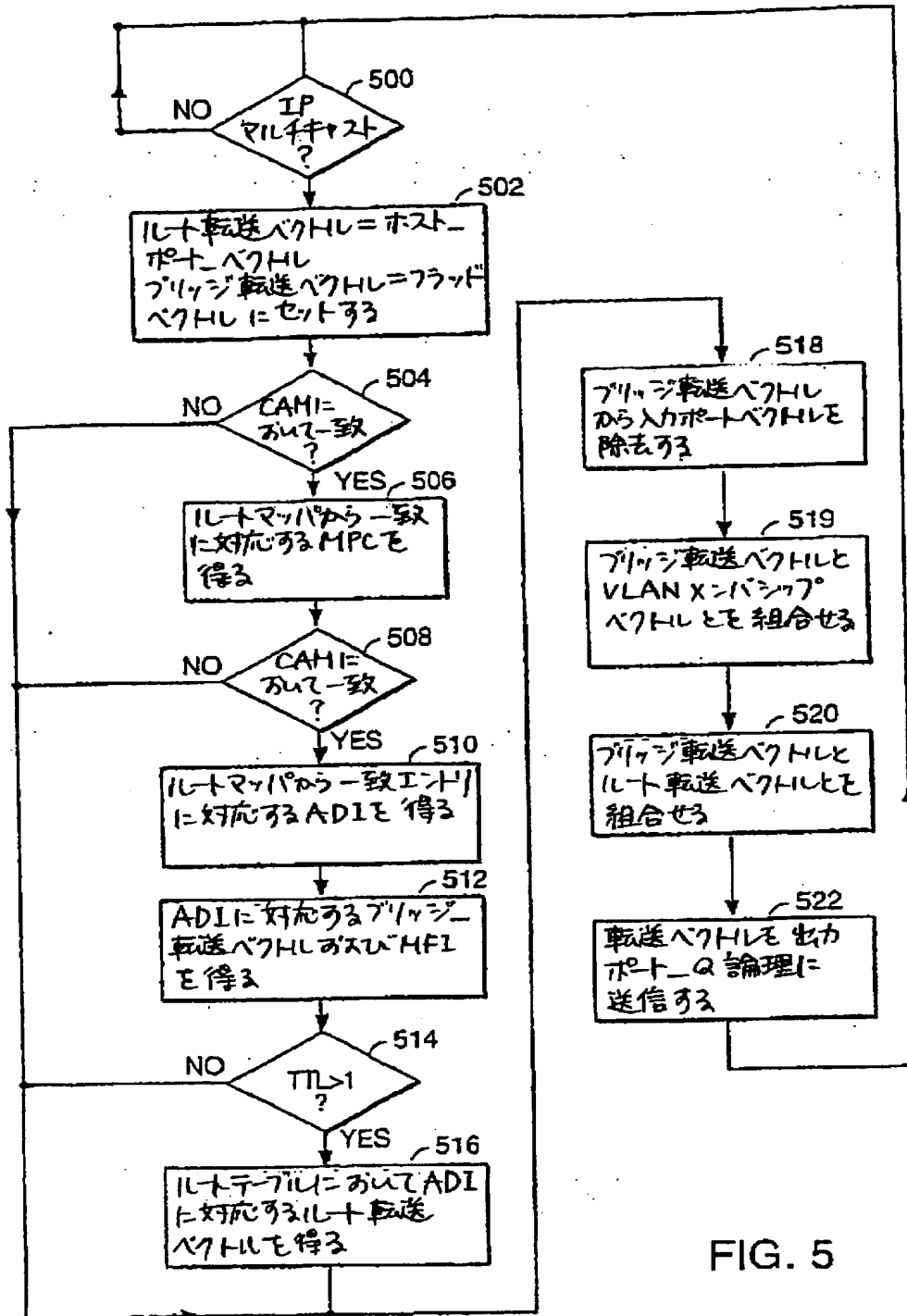


FIG. 5

【図4】

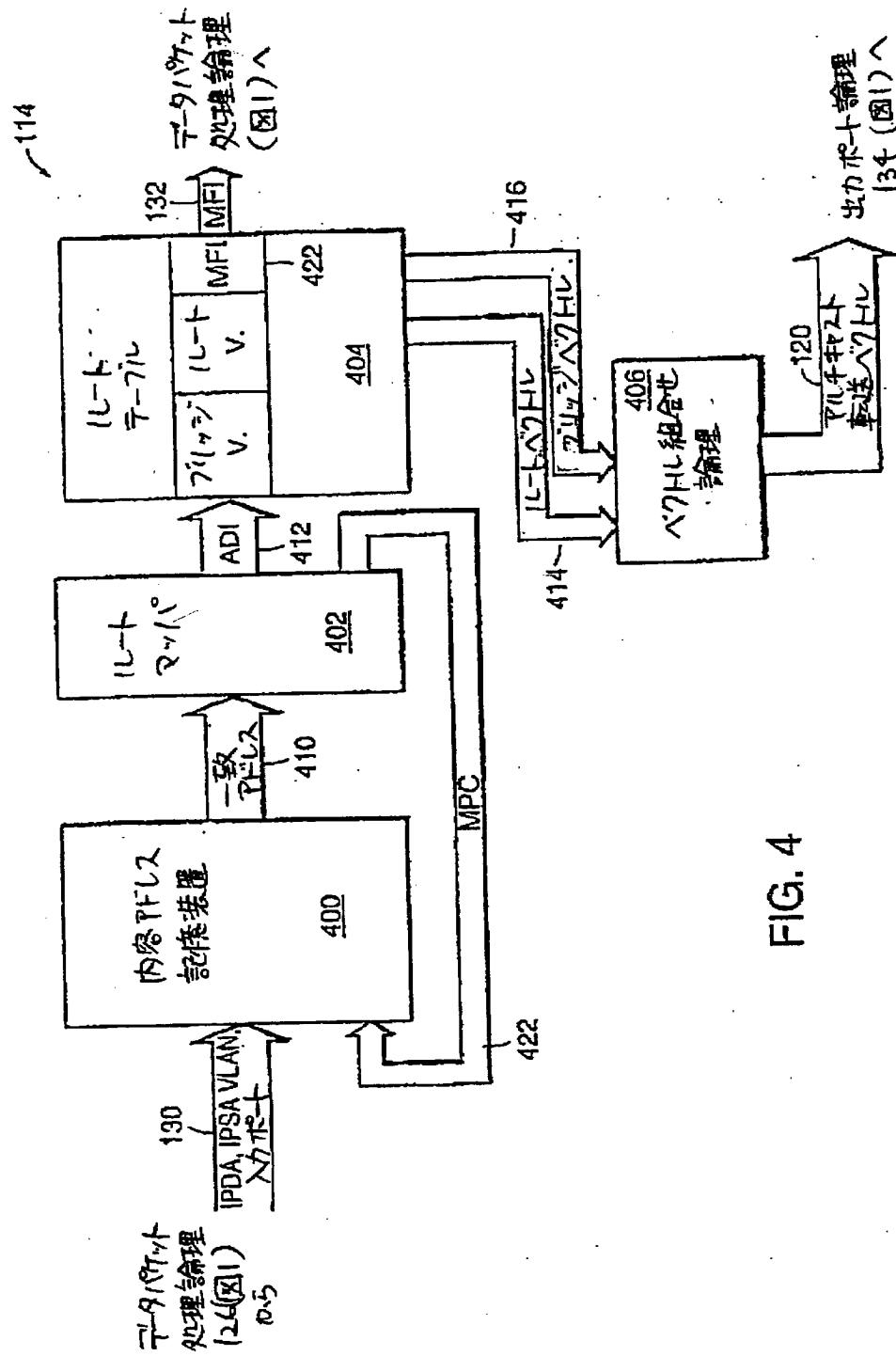


FIG. 4

【図6】

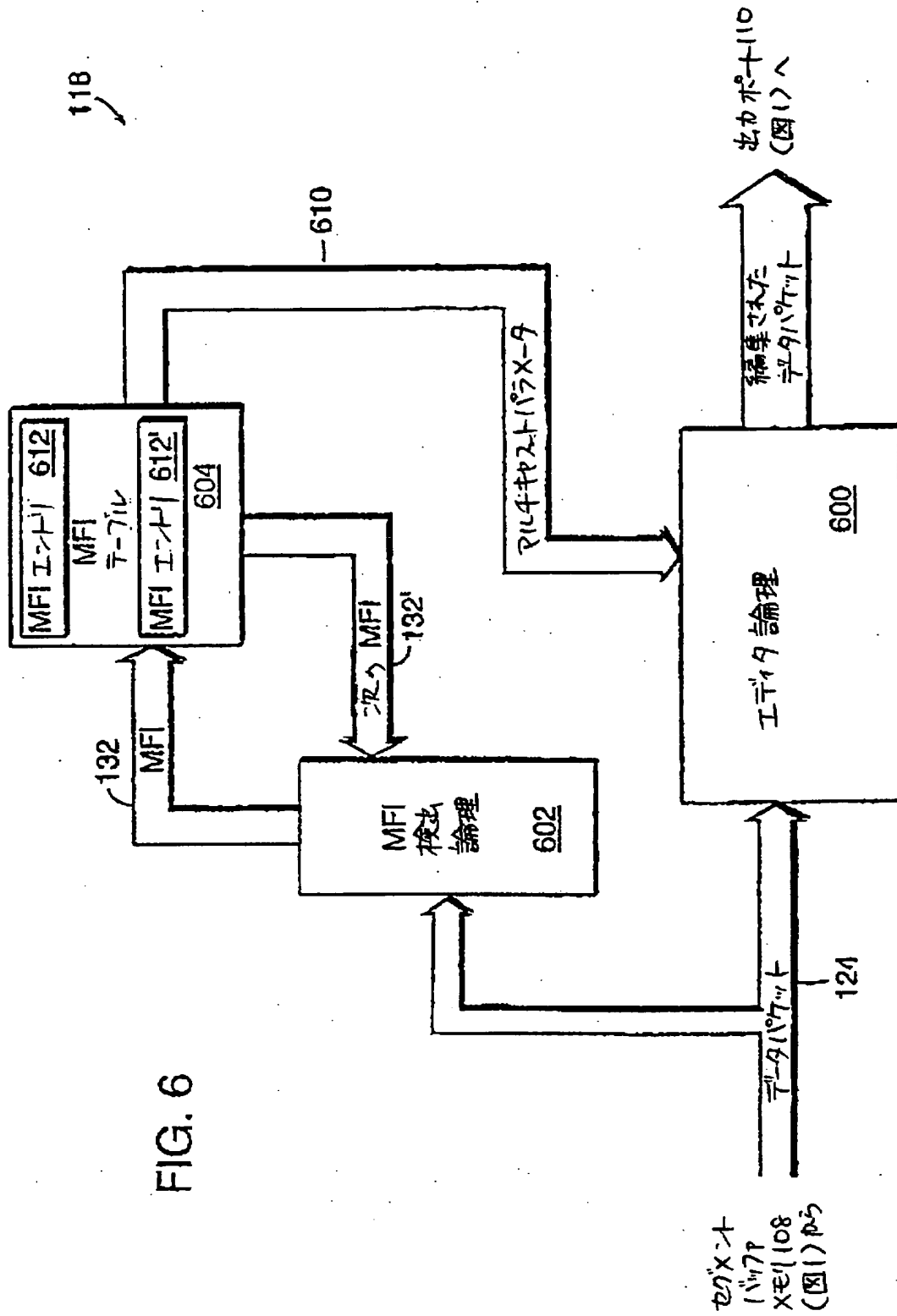
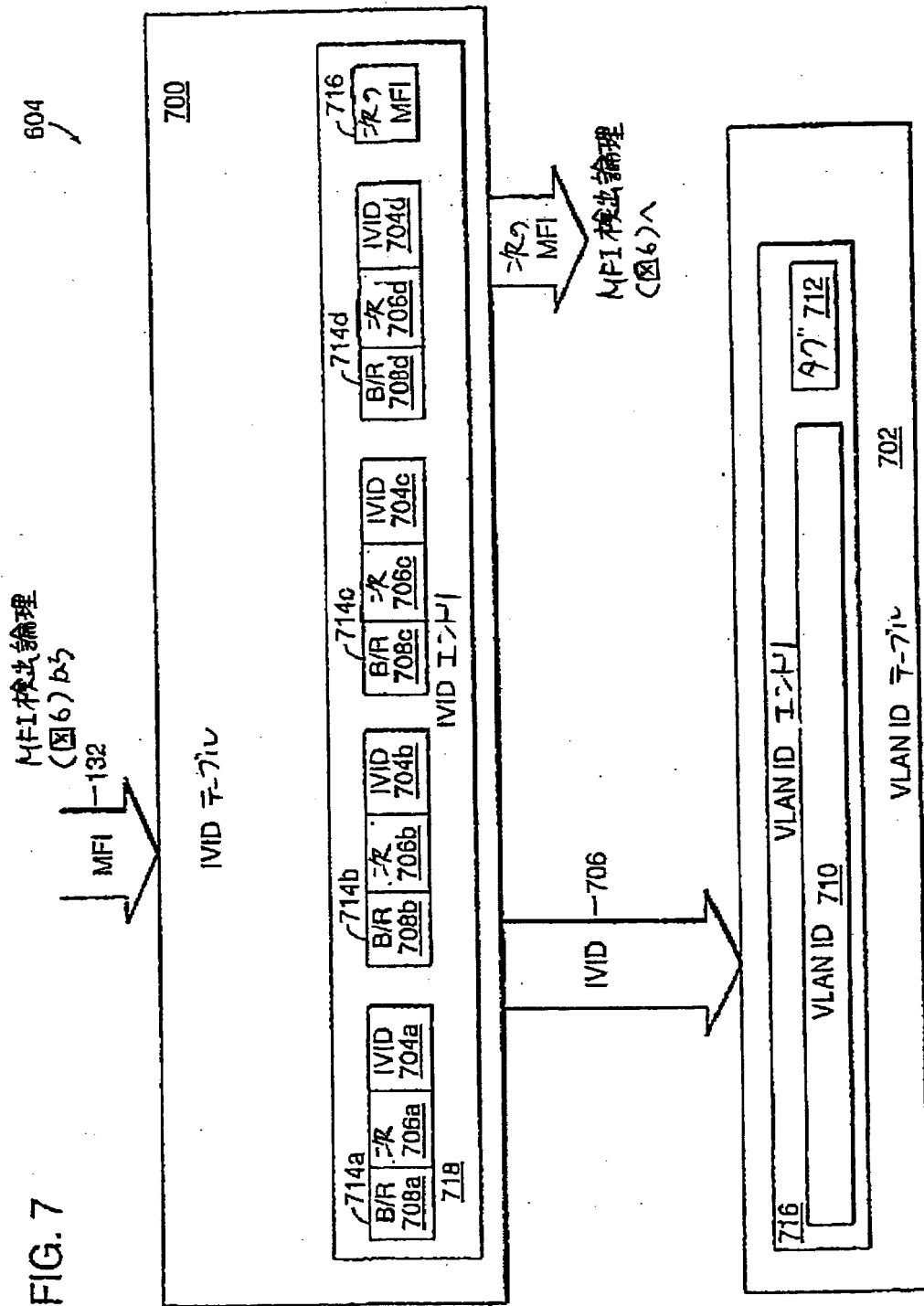
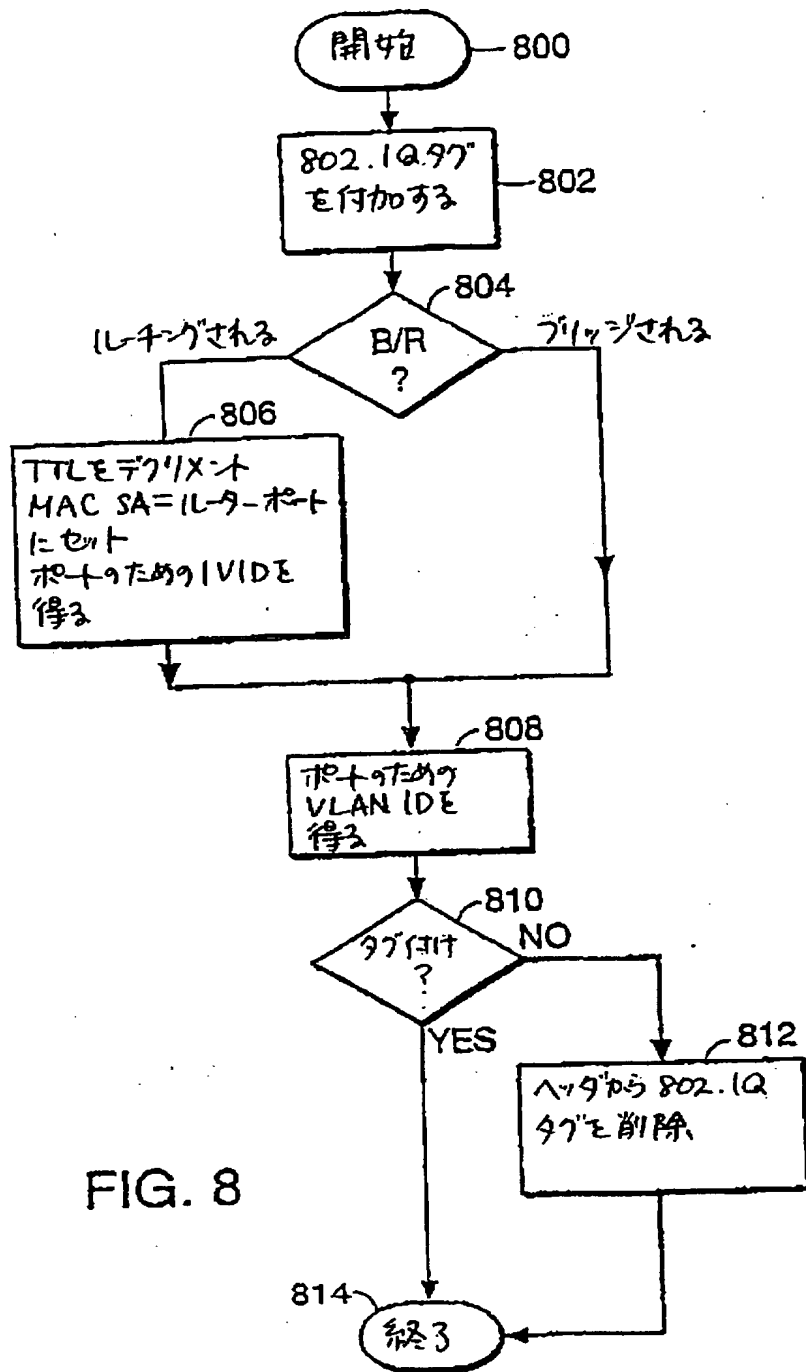


FIG. 6

【図7】



【図8】



【手続補正書】特許協力条約第34条補正の翻訳文提出書

【提出日】平成14年3月8日(2002. 3. 8)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】請求項1

【補正方法】変更

【補正の内容】

【請求項1】 マルチキャストグループのために受信されるマルチキャストデータパケットを記憶するメモリと、

出力ポート転送論理とを含み、出力ポート転送論理は、

マルチキャストグループのためのマルチキャスト転送エントリのリンクされたリストを含み、各マルチキャスト転送エントリは、マルチキャストグループのメンバに結合されるスイッチにおける各ポートのための修正エントリを含み、出力ポート転送論理はさらに、

修正エントリに依存してメモリから受信されるマルチキャストデータパケットのコピーを修正するエディタ論理を含む、スイッチ。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】請求項8

【補正方法】変更

【補正の内容】

【請求項8】 マルチキャストグループのために受信されるマルチキャストデータパケットをメモリに記憶するステップと、

出力ポート転送論理を設けるステップとを含み、出力ポート転送論理は、マルチキャストグループのためのマルチキャスト転送エントリのリンクされたリストを含み、各マルチキャスト転送エントリは、修正エントリに依存してメモリから受信されるマルチキャストデータパケットと関連付けられるマルチキャストグループのメンバである、スイッチ中の各ポートのための修正エントリを含み、さらに、

出力ポート転送論理において修正エントリに依存してメモリから受信されたマルチキャストデータパケットのコピーを修正するステップと、

出力ポート転送論理により、マルチキャストデータパケットの修正されたコピーを出力ポートへ、次の使用可能出力ポートサイクル時間において転送するステップとを含む、スイッチにおいてマルチキャストデータパケットを転送するための方法。

**【手続補正3】**

**【補正対象書類名】** 明細書

**【補正対象項目名】** 請求項15

**【補正方法】** 変更

**【補正の内容】**

**【請求項15】** マルチキャストグループのために受信されるマルチキャストデータパケットを記憶するメモリと、

出力ポート転送論理とを含み、出力ポート転送論理は、

マルチキャストグループのためのマルチキャスト転送エントリのリンクされたリストを含み、各マルチキャスト転送エントリは、マルチキャストグループのメンバーである、スイッチ中の各ポートのための修正エントリを含み、出力ポート転送論理はさらに、

修正エントリに依存してメモリから受信されたマルチキャストデータパケットのコピーを修正するための手段を含む、スイッチ。

## INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 H04L12/18 H04L29/06 H04L12/46

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)  
EPO-Internal, WPI Data, PAJ, IBM-TDB

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

C. DOCUMENTS CONSIDERED TO BE RELEVANT		Relevant to claim No.
Category *	Citation of document, with indication, where appropriate, of the relevant passages	
X	US 5 963 552 A (JOO SANG-WAN ET AL) 5 October 1999 (1999-10-05)	1-3, 8-10, 15-17
A	column 3, line 22 -column 6, line 37	4-7, 11-14, 18-21
X	US 5 898 687 A (HARRIMAN GUY ET AL) 27 April 1999 (1999-04-27)	1-3, 8-10, 15-17
A	abstract column 3, line 46 -column 5, line 36 column 7, line 56 - line 67	4-7, 11-14, 18-21

☒ Further documents are listed in the continuation of box C.

☒ Patient family members are listed in annex.

\* Special categories of cited documents:

- \*A document defining the general state of the art which is not considered to be of particular relevance
- \*E earlier document but published on or after the international filing date
- \*L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another claim or other special reason (as specified)
- \*O document referring to an oral disclosure, use, exhibition or other means
- \*P document published prior to the international filing date but later than the priority date claimed

- \*T later document published after the examination filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*C document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*V document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more others: such documents, such combination being obvious to a person skilled in the art.
- \*S document member of the same patent family

Date of the actual completion of the international search

4 July 2001

Date of mailing of the international search report

11/07/2001

Name and mailing address of the ISA  
European Patent Office, P.B. 5616 Patentaan 2  
NL - 2280 HV Rijswijk  
Tel (+31-70) 340-2040, Tx. 31 651 epo nl  
Fax (+31-70) 340-3016

Authorized officer

Peeters, D



## INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/CA 00/01420

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	XIONG Y ET AL: "Analysis of multicast ATM switching networks using CRWR scheme" COMPUTER NETWORKS AND ISDN SYSTEMS,NL,NORTH HOLLAND PUBLISHING. AMSTERDAM, vol. 30, no. 8, 1 May 1998 (1998-05-01), pages 835-854, XP004122815 ISSN: 0169-7552 page 836, paragraph 2 -page 837; figures 1,2	1-21
A	"MULTICAST/BROADCAST MECHANISM FOR A SHARED BUFFER PACKET SWITCH" IBM TECHNICAL DISCLOSURE BULLETIN,US,IBM CORP. NEW YORK, vol. 34, no. 10A, 1 March 1992 (1992-03-01), pages 464-465, XP000302372 ISSN: 0018-8689 the whole document	1-21

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/CA 00/01420

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5963552 A	05-10-1999	NONE	
US 5898687 A	27-04-1999	NONE	

## フロントページの続き

(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AP(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

**THIS PAGE BLANK (USPTO)**